DESIGN AND ANALYSIS OF EXPERIMENTS FOR MODEL DISCRIMINATION IN UNIRESPONSE AND MULTIRESPONSE SYSTEMS

A Thesis Submitted

In Partial Fulfilment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

SANTOKH SINGH

to the

DEPARTMENT OF MATHEMETICS

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

AUGUST, 1986

CERTIFICATE

Certified that this work, entitled, "DESIGN AND ANALYSIS OF EXPERIMENTS FOR MODEL DISCRIMINATION IN UNIRESPONSE AND MULTIRESPONSE SYSTEMS" by Santokh Singh, has been carried out under our supervision and has not been submitted elsewhere for a degree.

J. D. Borwankar
Professor
Department of Mathematics
Indian Institute of Technology
Kanpur, India

Musti S. Rao
Professor
Department of Chemical Engineering
Indian Institute of Technology
Kanpur, India

August 17, 1986.

CERTIFICATE

Certified that this work, entitled, "DESIGN AND ANALYSIS OF EXPERIMENTS FOR MODEL DISCRIMINATION IN UNIRESPONSE AND MULTIRESPONSE SYSTEMS" by Santokh Singh, has been carried out under our supervision and has not been submitted elsewhere for a degree.

J. D. Borwankar
Professor
Department of Mathematics
Indian Institute of Technology
Kanpur, India

Musti S. Rao
Professor
Department of Chemical Engineering
Indian Institute of Technology
Kanpur, India

August 17, 1986.

CERTIFICATE

This is to certify that Mr. Santokh Singh has satisfactor completed the course requirement for the Ph.D. program in Statistics. The courses credited by him include

M601 Graduate Mathematics I
- (i) Complex Analysis
- (ii) Probability and Statistics

M603 Graduate Mathematics III
- (i) Functional Analysis
- (ii) Topology
- (iii) Measure Theory
- (iv) Linear Algebra

M741 Advanced Estimation Theory

M542 Regression Analysis

M841 Topics in Statistics : Admissibility of Estimators

M592 Numerical Analysis

M502 Computer Programming.

Mr. Santokh Singh was admitted to the candidacy of the Ph.D. degree in August,1983, after he successfully completed the written and oral qualifying examinations.

Head
Department of Mathematics

Convenor
Departmental Post-Graduate Commit

Indian Institute of Technology
Kanpur.

ACKNOWLEDGEMENTS

I wish to acknowledge my indebtedness to Professor J.D. Borwankar of the Department of Mathematics and Professor M.S. Rao of the Department of Chemical Engineering at IIT Kanpur for their useful criticism and suggestions during the course of this work.

Acknowledgements are also due to other people who helped me directly or indirectly in the development of this work. I owe a great debt to my parents for their help and inspiration throughout my academic career. My wife Kuldip and son Vikram have been very patient throughout my Ph.D. program. I gratefu acknowledge their cooperation. It is a pleasure to put on records my thanks to Professor J.L. Batra of IME Program who has been a constant source of encouragement. Besides, a numbe of my friends have been of considerable assistance; Paramjit G is to be mentioned in particular.

I would also like to express my gratitude to Prof. J.B.Sh (former Head) and Prof. R.K. Jain (Head) of the Department of Mathematics for their help and cooperation. Thanks are also d to other members of the department for fruitful discussions I had with them at several occasions. The excellent job Mr. G.L. Misra has done in typing the manuscript deserves lot appreciation. I am also thankful to Mr. A.N. Upadhyaya for doi a neat job of cyclostyling.

Finally, I am greatly indebted to the Director, IIT Kanpu for allowing me registration in the Ph.D. program.

Santokh Singh

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SOME NOTATIONS AND SYMBOLS

Notations

ABS	absolute value
E	expectation operator
exp	exponential
$E^{(u)}$	the expectation under model u
sup	supremum
tr(A)	trace of the matrix A
t_r	r-variate t-distribution
Δ	criterion proposed in literature
$\emptyset$	the criterion proposed in this work

Symbols

(u)	(as superscript); under model u
$\hat{}$	(on top) maximum likelihood estimator (m.l.e)
$\sim$	(on top) some estimator
*	(on top) some estimator also used for some constants
=	(on top) some estimator
$\sim$	(at bottom) a vector
$\approx$	(at bottom) a two dimension array
Π	a statistical population
$\prod$	product
$\cap$	a process
$\|A\|$	determinant A
$\propto$	proportional
$\sum$	the summation
ℓ	log likelihood
$\mathcal{C}$	estimation criterion function
A'	transpose of A

SYNOPSIS

The significance of model building in a system study approach is well recognised. The requirement is often a mechanistic model, i.e., a set of mathematical equations which can describe the physical mechanism of the underlying process reasonably well. So far as arriving at the final form of an appropriate model is concerned, the whole exercise may not be as simple as one expects, for it may not simply be the estimation of the unknowns appearing in the equation(s) of the given model. The investigator frequently comes across situations where several models have been postulated for the same system. This asks for discrimination among the competing models and selection of the best, before one proceeds to the task of model fitting. Further, it may be noted that the occurrence of situations which involve more than one response is not uncommon. Besides, the models nonlinear in parameters cannot be debarred from competition.

For the present study we identify our problem as model discrimination in uniresponse and multiresponse systems. The cases of both the linear and nonlinear models have been planned to fall under the purview of this work. The task of model discrimination is normally accomplished sequentially. A procedure for this purpose, therefore, consists of the application of a discrimination criterion and, if required,

the use of a design criterion. Of the several procedures proposed in literature to that end, some have limitations while a few others suffer from certain drawbacks. For example, the methods proposed for bimodel problems cannot be applied to multimodel situations; the methods devised for discrimination among linear models may not work for nonlinear models; and a procedure meant for a uniresponse system can not handle the problem if the system happens to be of multiresponse nature. Furthermore, the drawbacks in some of the available procedures make one hesitant in using them. For example, the decision taken through the procedures which use posterior probability as a model adequacy criterion may not be reliable unless the observations from a large number of designed experiments have been utilized. Similarly, the methods which neglect the covariance structure of prediction errors may not do full justice to all the models, while the demanding nature of certain procedures sometimes makes their use restrictive. In Chapter 1, where we introduce the problem, the limitations and drawbacks of some procedures proposed in literature have been discussed in detail.

In the present study we attempt to resolve some of the difficulties which an investigator is likely to face in actual practice. Now, it may be borne in mind that the basic step in building a mechanistic model for a system is to establish the true nature of the underlying process. To that purpose it is important to appreciate that randomness is somehow imbued

into the process through one or the other sources, thus rendering it probabilistic in nature. The observations arising from a system as a result of certain inputs can, therefore, be presumed to be realizations of some random variable(s), $\underset{\sim}{Y}$, which may be characterized through a model in terms of the physical parameters of the process and the input variables. We, therefore, assume probability distributions characterizing the populations supposed to have been generated by the rival models. In a situation where several models have been proposed there would be a distribution due to the true model (hypothetical), $df^{(0)}$ (say), on the one side and a host of alternative distributions under the proposed models, $df^{(u)}$, $u = 1,2,\ldots,m$ (say), on the other. Based on the distributions $df^{(0)}$ and $df^{(u)}$ we develope an index, called Discrimination index, through which the claim that the proposed model, u, is the best alternative to the true model can be verified. This index, in fact, decides the relative adequacy of a model. If the available set of observations do not possess sufficient discriminatory power the above index may not show enough evidence in favour of a single model. In that case a few more observations acquired through experiments specific to the purpose of discrimination may prove to be helpful. This necessitates the use of a design criterion. The criterion that we propose in this work is based on the alternative probability distributions of the random variable(s), $\underset{\sim}{Y}_{n+1}$, on which a discriminatory observation is yet to be realized. This is, in fact, a weighted function involving all possible pairs of the

rival models. The weights employed in this function are formulated to meet the basic requirement that a pair consisting of distinct models receives less importance and the one with close rivals is allowed to play a greater role in designing an experiment. It may be noted that the variance-covariance structure of errors has been given due consideration in the formulation of both the design and discrimination criteria. It has been argued that the maximization of the proposed design criterion produces optimal setting of the experiment. All this has been done in Chapter 2. In this chapter itself we introduce a statistic for testing certain hypothesis relevant to the problem.

In Chapter 3, we subject the general design criterion to different sets of assumptions when the system is uniresponse. The criteria thus obtained are applicable to linear as well as to nonlinear models. Chapter 4 deals with multivariate linear and nonlinear models. Once again the design criterion is exposed to the multivariate analogues of the same sets of assumptions. The development of the different forms of the basic design criterion is based on the Bayesian approach. The proofs of some results used in these two chapters for arriving at the final forms of the criteria have been given in the Appendices. Before the formulae developed in Chapters 3 and 4 can be applied to a specific problem one needs estimates of certain unknowns, such as parameters of the models and the variance-covariance

structure of errors. Some theory of estimation to that effect has been given in Chapter 5 from where a suitable estimation criterion or an appropriate formula can be chosen according to the assumptions applicable in a given situation. In the last chapter, i.e., Chapter 6, we demonstrate the implementation of our procedure comprising of design and analysis for model discrimination. This is done through Monte Carlo simulation. The results thus obtained are compared with those reported elsewhere. Through application to varried problems, the proposed procedure has been found to converge faster to the true model. Besides, the emphasis is always on the best model followed by less apt ones rather than on bad models. The discrimination thus achieved looks pretty sharp. The index which is employed for assessing the discrimination shows a consistent trend and does not oscillate as the posterior probability does. The proposed weights, too, have been seen to be appropriate.

CHAPTER 1

INTRODUCTION AND REVIEW OF THE PREVIOUS WORK

1.1 SYSTEMS AND MODELS

In the present day world 'system' is perhaps one of the most widely used concept in scientific investigations. The problems of mechanics are solved conveniently after identifying a mechanical system; the biologist thinks of an organism as a biological system; the chemical engineer studies a reaction through a chemical process control system; an economist, studying the economy of a state, has in his mind an economic system; so on and so forth. The system study seems to be the way of investigation in almost all the disciplines. In whichever field one may be using this concept, the term system has a common underlying meaning. In fact, one can think of a system as a unit consisting of interdependent elements or components which interact regularly so as to complete an implicitly or explicitly assigned job. This is the world inside the system. But, there is a world outside, too, which exercises its influence through the input. Whenever the system receives an input the process operating in the system transforms it into an output through interaction of its elements. If the underlying system is deterministic, a particular input will always result into the same output. On the other hand, a probabilistic system

may respond with any one from amongst a range or distribution of outputs. It can, however, be argued that observations or measurements on the output of a system are random rather than deterministic. In certain cases randomness is inherent in the system, while in others only certain manifestations are observed because of insufficient information about the system response(s) or a lack of techniques to observe the output. Often, the observer is just negligent or careless. In general, systems can, therefore, be treated stochastically.

Unfortunately, the behaviour of a system is not always exemplary : a biological system may be disturbed by a disease; a chemical system is subject to catalyst decay; an ecological system is likely to be upset by pollution; and an economic system may be interrupted by inflation; etc. One must, therefore, attempt to save a system from external disturbances so as to improve the quality of its performance. There are several ways to do this. One can build a new system and discard the old one. But, when such a replacement is not possible, as is generally the case, the only alternative is to live with the existing system. Efforts can, however, be made to provide it with proper care and exercise supervision. Whatsoever be the case, the original problem always arises in the real world; sometimes in the controlled conditions of a laboratory and sometimes in the much less understood environment of everyday life. In any case, a system study, aimed at an effecient control of the system, is always concerned with the process (a series of events) which

generates one or more responses as a result of certain input. The probable future course of events must be predicted in order to evolve ways for imposing forces on the pertinent system so that it may be moved into a direction deemed to be desirable. The use of the actual process for this purpose may not be feasible in certain situations and not desirable in others for economic reasons. One may, therefore, resort to modelling. In ordinary language the word 'model' has many meanings. But the model which we use in this study is composed of algebraic equations rather than being a miniature of the system or something else. Such a model would, no doubt, be a mere mathematical representation of the underlying mechanism of the process, but could still serve the purpose with reasonable faithfulness. In fact, this provides us with the most powerful analytical tool for studying and investigating a system. The objectives, such as, control of the process and prediction or optimization of the output can be conveniently achieved through a mathematical model. It is for these reasons that almost all the disciplines which adopt a 'system approach' also use the concept of model as a basis for developing solutions to various relevant problems.

1.2 THE MODELLING PROCESS

A model may be a mechanistic representation of the system or just an empirical relation, depending on the purpose of an

investigation. In a given situation, if the aim is to predict or optimize, an empirical model might be entirely satisfactory. On the other hand, if the primary interest is not merely to predict the response(s) over a limited region, but rather to elucidate the mechanism or to obtain meaningful results in regions relatively unknown then a mechanistic representation of the process may be more appropriate. In fact, a mechanistic model can provide useful information where the optimization of the design of a system leans heavily on what is actually taking place. But, unfortunately, the model, even if it is a mechanistic one, is seldom exact. One must, therefore, satisfy onself with a close approximation so long as it is capable of simulating the mechanism and, thus, represent the real world phenomenon to a reasonable extent.

The utility of a mathematical model to the scientists and engineers is immense, in that, it is a useful tool for compressing large amount of information, providing new insight into the process, and suggesting ways for the future development of the system. The process of modelling, therefore, is an important operation and ought to be carried out in stages, crossing each stage carefully. While one has to be very particular in formulating the requirements of a model, one must also be fussy in hypothesizing its form. If there happens to be a single model considered appropriate for describing the process of the system, the only job of the investigator is to

secure precise estimates of the parameters, i.e., the unkonwns appearing in the model equation(s). The problem is of estimation . But, things may not always be that simple. One might find oneself at a point where "all roads seem to lead to Rome". That is to say, there may be several models capable of explaining the mechanism of the system. In such situations, a model can be actually put to use only if a single model is picked up out of the given lot and fitted precisely through the available data. The problem, now, is of discrimination among the equally plausible models. The investigator must take a decision on the right path to be followed. Once the most appropriate model has been selected from amongst the given lot, one may start brushing up the estimates of its parameters. Discrimination may either be possible with whatever data are available at hand in the beginning or one may require additional observations. In the latter case, it is naturally desirable to have such observations as can help in a fast and sharp discrimination. This needs designing additional discriminatory settings of the input variable(s). Thus one starts with a discrimination criterion but may require a discriminatory design criterion if the given set of observations do not prove to be strong enough to pick up the best model.

1.3 MATHEMATIZATION OF A PROCESS

Quite often, the experimenter deals with such situations where only one characteristic of the output is being observed either because the output is unidimensional or the interest of the experimenter centers around one aspect of the system only. For example, in an experiment, one might be interested only in the tar content of a gas stream as a result of input of gas inlet temperature and rotor speed. But, situations are not uncommon when observations from the process come as measurements on more than one characteristic of the system. For example, in a study of a chemical process, with each input, one might observe yield and density of the product.

Consider a system whose mechanism is made up of p parameters, $\theta_1, \theta_2, \ldots, \theta_p$, and suppose that it responds in terms of the r dependent variables, $Y_1, Y_2, \ldots, Y_r$, as a result of certain input of the q independent variables, $\xi_1, \xi_2, \ldots, \xi_q$. Mathematically, such a system can be represented by a set of r equations

$$E(Y_i) = \eta_i (\xi_1, \xi_2, \ldots, \xi_q ; \theta_1, \theta_2, \ldots, \theta_p),$$

$$i = 1, 2, \ldots, r, \qquad (1.3.1)$$

where $r = 1$ or $r \geq 2$, according as the system is uniresponse or multiresponse and the function η_i can be linear or nonlinear in parameters, $\theta_1, \theta_2, \ldots, \theta_p$. In fact, (1.3.1) comprises the true mathematical model of the given system, but is no more than a mere hypothesis. One, therefore, seeks a close

mathematical representation. Specific to this study, we consider a situation in which not one but a number of physically meaningful models have been postulated. To start with, these models are thus assumed to describe the process equally well. The given system can, therefore, be alternatively represented by any one of the m(say) rival models :

$$E^{(u)}(Y_i) = \eta_i^{(u)}(\xi_1,\xi_2,\ldots,\xi_q;\ \theta_1^{(u)},\theta_2^{(u)},\ldots,\theta_{p_u}^{(u)}),$$

$$i = 1,2,\ldots,r;\quad u = 1,2,\ldots,m, \qquad (1.3.2)$$

The modelling process in this case consists of, first, identifying 'the' model and then proceeding to improve upon the estimates of its parameters, if required.

1.4 REVIEW OF THE PREVIOUS WORK IN MODEL DISCRIMINATION

Which of the m models best represents the system? This has been one of the fundamental questions confronting researchers engaged in building models for various types of systems. Discrimination among the given models may provide an answer. In fact, this problem of selection of the most appropriate model from amongst the proposed ones has been widely discussed in the literature. Reilly and Blau (1974), Hill (1978), Singh and Rao (1981), Iyengar and Rao (1983, 1984), and several others have given extensive reviews of model discrimination procedures. The entire work in this field can be put into two broad categories, namely, the Bayesian and the Non-Bayesian. However, the various procedures proposed in the literature for discrimination among

rival models have one thing in common : almost all of them use the concept of divergence in one or the other way.

1.4.1 Review of the Previous Work : Uniresponse Case

Bimodel Case : Hunter and Reiner (1965) are, perhaps, the pioneers in demonstrating the use of divergence in designing experiments for model discrimination. In order to discriminate between two regression models they assume that the observations are normally distributed about zero mean with a constant variance, σ^2. According to them, an experimental run, (n+1)th(say), will result into a discriminatory observation if it is conducted at a setting of the input variable(s), $\underset{\sim}{\xi}$, where the two fitted surfaces are farthest appart. But, in the process of optimization this means an immense effort involved in computing estimates of the model parameters (especially, in case of nonlinear models) for every choice of $\underset{\sim}{\xi}_{n+1}$ in its operability region. Hunter and Reiner have, however, resolved this difficulty by proposing an approximate design criterion,

$$\Delta_1(\underset{\sim}{\xi}_{n+1}) = \max_{\underset{\sim}{\xi}_{n+1}} \{\eta^{(1)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(1)}) - \eta^{(2)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(2)})\}^2, \quad (1.4.1)$$

where the divergence between the responses, expected at the (n+1)th run, is always evaluated through estimates, $\hat{\underset{\sim}{\theta}}_n^{(1)}$, $\hat{\underset{\sim}{\theta}}_n^{(2)}$, of model parameters, based on the available n observations. The experimental points are recommended to be designed, sequentially, by solving the equation (1.4.1) at each stage till the adequacy of one of the models is confirmed.

Although, developed through the assumption of the truth of one of the models, taken alternatively, the criterion in its final form does not require any such assumption. This, in fact, is the result of combining the two noncentrality parameters into one function which, therefore, becomes symmetric in terms of models. In actual practice, the investigator is expected to confirm in the beginning if both models are adequate, neither model is adequate, or one model is adequate. It is only in the last situation that designing of additional experiments is required till it is determined as to which of the two models is adequate. As regards the testing of adequacy, the authors have proposed the use of F-test, though on heuristic basis only; theoretically, the application of such a test is not valid. The replications needed to obtain a valid estimate of σ^2, as is required for this test, are claimed to occur naturally through the proposed design criterion. The authors have, however, suggested to deviate a designed point, slightly, if it could become a replicate.

The criteria which Atkinson and Fedorov (1975a) used for designing discriminatory experiments in a bimodel situation are also based on the noncentrality parameter, $\lambda_2(\zeta)$, (i.e. the sum of squares for lack of fit) of the second model, after it has been assumed that the first model is true and that its parameters are known. The sequential and nonsequential design criteria, actually proposed by them with different types of requirements, have been basically developed through the equation

$$\Delta_2(\zeta^*) = \sup_{\zeta} \lambda_2(\zeta) \quad , \qquad (1.4.2)$$

where ζ^* is the resulting design, and ζ is a normed measure defined on the compact design space. The solution of this equation will result into what they have termed as locally optimal design. In an equivalence theorem they have laid down the necessary and sufficient conditions for a design to be T-optimum as well as listed some of the properties of such a design. This theorem is, therefore, useful in confirming the optimality of a given design or for investigating a proposed numerical procedure for the construction of a discriminatory design. Atkinson and Fedorov (1975a) have, in fact, established through this theorem that the procedure of Hunter and Reiner (1965) produces designs which are asymptotically T-optimum.

Nevertheless, the optimal designs of Atkinson and Fedorov (1975a) greatly depend on the information as to which of the two models is true as well as on the values used for the parameters of this model. But, if one already knows about the true model to that extent the whole exercise of construction becomes redundant. The alternative approaches suggested for getting rid of this restriction, too, may not prove to be fruitful as they further lead to other requirements such as, linearity of the models and prior information about the rival models and their parameter distributions. Thus one may find the use of their nonsequential method restrictive and difficult. The practical utility of their work mainly lies in their sequential

procedure. This procedure consists of starting with an initial design, estimating the parameters of both the models through the least squares method, and then taking the next experiment at a point where the divergence between the two expected responses is maximum. The designing of experiments is continued till a test of adequacy confirms the suitability of one of the two rival models. This leads to a design which, no doubt, would be T-optimum but has, nevertheless, certain shortcomings. According to them, the design thus realized would be data dependent and would require a considerable number of trials to wipe out the effect of a large error in observations. Furthermore, the procedure may lead to a singular design. Thus in order to construct an efficient design it is desirable that the observations be as free from experimental error as possible. Besides, the investigator must start with a nonsingular design. Added to all these is the requirement of appreciable replication for testing adequacy of the models. Nevertheless, the availability of the desired replications, too, may not work if one model happens to be a degenerate form of the other. In this case the inclusion of another suitable model may be a way out. But, this might further complicate the task of discrimination, in that, it leads to a multimodel discrimination problem.

Multimodel Case : The use of both the criteria discussed so far is limited to bimodel problems only, while in practice the investigator might have to choose from a set of more than two models. Using rather a Bayesian approach Roth (1965) has

developed a criterion, based on Hunter and Reiner's idea only, which could be used in a multimodel problem. This criterion is a weighted average of the total separation among m rival models; the weights being the prior probabilities of models. The (n+1)th experimental point is chosen at the maximum of

$$\Delta_3(\underset{\sim}{\xi}_{n+1}) = \sum_{u=1}^{m} \left[P_n^{(u)} \prod_{\substack{v=1 \\ v \neq u}}^{m} \left| \eta^{(v)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(v)}) - \eta^{(u)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(u)}) \right| \right], \tag{1.4.3}$$

where $P_n^{(u)}$ is the prior probability and $\hat{\underset{\sim}{\theta}}_n^{(u)}$ are the estimates of parameters of model u, based on the previous n observations. After each sequential run the probabilities, used as weights in the criterion function, are updated through the Bayes' formula and used for decision making. Reilly (1970) has criticized Roth's criterion on the ground that it fails to consider uncertainty associated with prediction. It may be argued that it is not safe to use this criterion if the parameters of all the models are not estimated with the same precision. In fact, large divergences due to imprecise estimates of certain model parameters might unnecessarily inflate the criterion value. In this case, the effect of separations resulting from precisely estimated responses is subdued. Since the criterion does not take into account the magnitude of the precision of the predicted responses this may result into a superficial maximum and consequently into a wrong design. Another shortcoming of this criterion lies in its slow convergence to the desired discriminatory level, when used for discrimination among mathematically similar models.

Froment (1975) has proposed a straight forward extension of the Hunter-Reiner criterion to multimodel situations. This consists of maximizing

$$\Delta_u(\underset{\sim}{\xi}_{n+1}) = \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} \{\eta^{(u)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(u)}) - \eta^{(v)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(v)})\}^2 \quad (1.4.4)$$

with respect to $\underset{\sim}{\xi}_{n+1}$, the (n+1)th setting of the input variable(s). Hosten and Froment (1976) have, instead, preferred the absolute value of the difference between the responses predicted through models u and v. Discrimination in both the methods is done by screening out bad models, sequentially. So far as the detection of ill-fitting models is concerned they use the fact that the mean sum of squares for lack of fit of a model will be an unbiased estimate of the error variance for the correct model only. Thus, if σ^2 is known or an estimate of σ^2 is available, the inadequate models are discarded at any stage through an F-test. In case of complete ignorance about the magnitude of variance of errors this job is proposed to be done by means of Bartlett's χ^2 statistic. This statistic, basically, tests the homogeneity of the estimates of σ^2, obtained through the rival models. At any stage the model with the largest estimate of σ^2 is dropped if χ^2 exceeds the tabulated value. This way more and more models are eliminated as discrimination advances from one stage to another. The application of this statistic requires each model to be at least locally linear. Besides, χ^2 being sensitive to departures of data from normality, care has to be taken with outliers. So far as the independence of estimates of

variance is concerned, Froment (1975) has observed that the procedure is not sensitive to it. Domez and Froment (1976) have used $\Delta_4(\underset{\sim}{\xi}_{n+1})$ as the design criterion function for discriminating among fifteen rate equations of dehydrogenation of 1-Butane to Butadiene on a chromium-aluminium-oxide catalyst in a differential reactor. The technique worked successfully in picking up one equation as representative of the system under consideration.

Another extension of a bimodel discrimination method to the case of several models is proposed by Atkinson and Fedorov (1975b). Once again, they utilize the concept of maximizing the noncentrality parameter of a model. In case of m rivals this requires simultaneous maximization of (m-1) noncentrality parameters, assuming that one of the models is true with known parameters. To that purpose they choose to adopt a maximin approach in solving a continuous design problem and title the designs thus obtained as T-optimum. Like its bimodel analogue, an equivalence theorem prescribes the necessary and sufficient conditions for a design to be optimal. This theorem is useful in determining the optimality of a given design but has little use in constructing an optimal design. The authors have, however, proposed sequential procedures for actually realizing such a design, mainly, in three types of situations. If there is one closest rival, as can be seen by comparing the residual sum of squares, the new experiment is recommended to be chosen in such a way that the divergence between the responses predicted through the best fitting models is maximum. The sequential method of Atkinson and Fedorov

may produce a nonoptimal design if there happens to be more than one equally close models. If there are two closest rivals, the situation can still be managed through the iterative procedure proposed for this case. But the case of several equidistant models is difficult to handle.

Atkinson and Fedorov (1975b) have also proposed non-sequential methods for constructing an optimal design under certain assumptions which are not different from the ones required in their bimodel analogues. The occurrence of equality of two or more noncentrality parameters may once again add to the difficulty which the investigator may already be facing in meeting some of the basic requirements such as, linearity of rival models and complete knolwedge about the true model or else prior information in the form of the probability distributions and prior probabilities of the models being discriminated.

We now discuss those criteria which are based on the argument that in discriminating among models the point of maximum divergence between the expected responses may not necessarily be the point of maximum discrimination. In fact, some researchers in this field have felt that it is the divergence in responses of models relative to the limits of their errors which plays a more important role in generating a good discriminatory point. According to them, the variances of the estimated responses must, therefore, come into picture. These researchers have actually taken note of this fact in the development of the design criterion.

For example, Fedorov and Pazman (1968) have included, in addition, the variances of the predicted responses. Assuming that the variance of the observations, $\sigma^2(\underset{\sim}{\xi}_{n+1})$, is a known function of the independent variable(s) and that the errors are normally distributed about zero, they propose to maximize

$$\Delta_5(\underset{\sim}{\xi}_{n+1}) = \left[\frac{1}{\sigma_u^2(\underset{\sim}{\xi}_{n+1}) + \sigma^2(\underset{\sim}{\xi}_{n+1})}\right] \left[\{\eta^{(u)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(u)}) - \eta^{(v)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(u)})\}^2 + \{\sigma_u^2(\underset{\sim}{\xi}_{n+1}) - \sigma_v^2(\underset{\sim}{\xi}_{n+1})\}\right] \qquad (1.4.5)$$

under the null hypothesis that model u is the true (u=1,2) one, where $\sigma_u^2(\underset{\sim}{\xi}_{n+1})$ is the variance associated with the prediction, $\eta^{(u)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(u)})$. In the absence of any knowledge about the true model they suggest the use of the smaller of the criterion values or some weighted combination of the two. The criterion function (1.4.5) seems to have originated from an earlier work of Pazman and Fedorov (1968), where the null hypothesis is about the unbiasedness of one of the two sets of estimates of the parameters of a given model rather than the truth of one of the two models. The criterion given by equation (1.4.5) has certain limitations. First of all, it can not be used if there are more than two competing models. Secondly, the requirement of the variance of observations to be a known function of the independent variable(s) can not be met so easily in actual practice.

The criteria discussed so far have been seen either neglecting the prediction variances or if they do not, they are

applicable to bimodel problems only. Box and Hill (1967) have proposed a criterion which not only takes into account the prediction errors but can also be used in multimodel situations. Besides, it does not require one of the rivals to be known in advance as a true model. This criterion, which utilizes the Shannon's (1948) concept of entropy, has been basically developed through the maximization of the change in entropy, R(say), expected due to the (n+1)th experiment yet to be conducted. Box and Hill have, however, used its upper bound, D (say), for designing a new experimental setting in a multimodel problem. In particular, for discriminating among m models they have assumed that under model u the errors are normally distributed about zero mean with a variance, σ^2 (known and constant for all experiments); that the models are linear or approximately linear in the neighbourhood of certain estimates of their parameters ; and that a locally uniform prior (non-informative) distribution is appropriate for the parameters of each model. With these assumptions they have obtained D to be of the form

$$\Delta_6(\underset{\sim}{\xi}_{n+1}) = \sum_{u=1}^{m-1}\sum_{v=u+1}^{m} P_n^{(u)} P_n^{(v)} \left[\frac{\sigma_u^2 - \sigma_v^2}{(\sigma^2+\sigma_u^2)(\sigma^2+\sigma_v^2)} + \right.$$

$$\{\eta^{(u)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(u)}) - \eta^{(v)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(v)})\}^2$$

$$\left.\left(\frac{1}{\sigma^2+\sigma_u^2} + \frac{1}{\sigma^2+\sigma_v^2}\right)\right], \qquad (1.4.6)$$

where $P_n^{(u)}$ is the probability of model u after n observations,

$\hat{\underset{\sim}{\theta}}_n^{(u)}$ is the maximum likelihood estimate (m.l.e) of $\underset{\sim}{\theta}^{(u)}$ based on the available n observations, and σ_u^2 is the variance of $\eta^{(u)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(u)})$. In order to decide upon the input for the (n+1)th exponential run, the criterion function $\Delta_6(\underset{\sim}{\xi}_{n+1})$ is maximized with respect to $\underset{\sim}{\xi}_{n+1}$. This is followed by a revision of the model probabilities $P_k^{(u)}$, by means of the Bayes' formula

$$P_k^{(u)} = \frac{P_{k-1}^{(u)} f^{(u)}(y_k)}{\sum_{v=1}^{m} P_{k-1}^{(v)} f^{(v)}(y_k)}, \quad k = n+1, \ldots; \; u = 1,2,\ldots,m, \tag{1.4.7}$$

where $f^{(u)}(y_k)$ is the probability density function (p.d.f.) of Y_k under model u and $P_{k-1}^{(u)}$ is the probability of model u at the previous, (k-1)th, stage. The quantity $P_k^{(u)}$ is utilized in assessing the status of each model at every stage in the sequential process which is terminated as soon as one of the given models emerges as the best model, i.e., the model with the highest posterior probability, $P_k^{(u)}$ (close to 1.0).

The Box-Hill procedure has been actually used by several investigators, but all of them have reported varied experiences with its implementation. Hunter and Mezaki (1967) have successfully applied this procedure to some practical problems in chemical engineering, while Froment and Mezaki (1970) and some other investigators faced certain difficulties in using it. The procedure has been subjected to considerable criticism which has

led to some modifications. Meter et al. (1970) and Reilly (1970) have expressed their surprise over the criterion itself; the maximization of the upper bound of expected entropy change, $\mathfrak{D}$, for realizing a design point looked strange to them. Reilly (1970) and Fedorov (1972) have instead obtained approximations to R. However, the examples in which these approximations have been used as a criterion function do not exhibit any significant difference in designs. Box and Hill (1967) have employed weights in the design criterion to ensure more emphasis on pairs of models with high probabilities, but Meter et al. (1970) are doubtful if these weights really do the job. In fact, if a term inside the square brackets in Δ_6 dominates the probability weightings, the criterion may select experiments carrying large amount of information in support of the models which are already on the verge of being dropped out. Buzzi and Forzatti (1983) suspect the criterion to pick up those points where the discrepency between variances of estimated responses is large, even if the divergence between the two predictions is small or negligible. It has also been observed that the design point obtained at the maximum of D may not be the one which maximizes R, although the proposed criterion is claimed to be based on the maximization of the expected entropy change. Furthermore, in the development of the criterion function, Box and Hill have assumed the ratio of each pair of prior densities to be unity, while Atkinson and Cox (1974) feel that this should make sense only if the parameters in both the models have the

same number and similar interpretations. Box and Henson (1970) seem to have this fact in mind when they expected an improvement in the formula (1.4.7) through the use of an alternative expression for the density $f^{(u)}$; namely,

$$f^{(u)}(\underset{\sim}{y}_{n+1}) = \exp\{-\frac{1}{2\sigma^2}(-s^{(u)} + p_u\sigma^2)\}\, c^{(u)}, \qquad (1.4.8)$$

where $s^{(u)}$ is the residual sum of squares from (n+1) observations, under model u which contains p_u parameters, and $c^{(u)}$ is the coefficient that accounts for the parameter indetermination. Buzzi and Forzatti (1983) have instead noted that even if this modification is introduced in the Bayes' formula, the posterior probability as a measure of the quality of a model remains equivocal. Their experience indicates that a change in the order of the designed experiments may change the inference on model selection. The numerical problems in the calculation of model probabilities through the modified formula arise partially due to the exponential relation between $f^{(u)}$ and $s^{(u)}$ and partially because of the form of $c^{(u)}$.

There are still more serious allegations on the Box-Hill procedure for model discrimination. Some researchers have even gone to the extent of remarking that this procedure can sometimes lead to false conclusions. Andrews (1971) fears that the procedure may even put a true model with more parameters in danger if a simpler model is close to it. In a simple problem of discriminating between two models, Atkinson and Cox (1974) have also observed that the posterior probability,

as a model adequacy criterion, tends to prefer a model with fewer number of parameters, although from significance testing point of view both the models have been found to be adequate.

Atkinson (1978), too, has pointed out that a simple Bayesian formula for the posterior probability of a model is misleading unless all the models have the same number of parameters. According to him, this formula leads to arbitrary inferences. He has, therefore, suggested an alternative quantity for choosing a model. Atkinson has further pointed out that the Box-Hill approach tends to force a choice between models ; even if both the models fit badly, the better fitting model will eventually be chosen. In fact, the probability of such a model can be brought as close to unity as desired, by continuing experimentation long enough. It is, therefore, important to check the adequacy of the chosen model through a suitable method before it is finally selected for actual use. In case of nested models, even though all of them may be true, the behaviour of the posterior probability does not indicate this reality. Box and Hill (1967) have considered an example in which all the rival models are generalizations of the model assumed to be true and the criterion picks up the simplest model after 15 points have been designed. Siddik (1972) has further investigated this situation and found that the probability of the simplest model keeps oscillating between 0.85 and 0.95, instead of converging to 1.0. Froment and Mezaki (1970),

Wentziemer (1970), and Hill (1976) have also observed the posterior probabilities of models to be oscillating considerably from one experimental run to another and have warned that the criterion must be used cautiously.

Finally, one notices that Box and Hill (1967) have, no doubt, catered for the variances of the estimated responses in the discriminatory design criterion, but have, at the same time, introduced a quantity, σ^2, the error variance, which is seldom known in practice. Froment and Mezaki (1970) have actually found this requirement a demanding one and instead used a precise estimate of σ^2. But, such an alternative may not always be possible, especially when replications are not available. Hill and Hunter (1969) have suggested a modification of the Box-Hill criterion itself so as to bring the case of unknown variance under its purview. In order to arrive at the final form of the criterion function they have used approximations to the functions involved in the integral which is supposed to yield an upper bound, D, of the expected entropy change in the unknown variance case. Their criterion is, thus, an approximation to D. The objection may, therefore, be raised to its accuracy. The criterion, however,looks quite similar to the Box-Hill criterion. On the other hand, Hosten and Froment (1976) have criticized the inclusion of the variances of predictions which include σ^2 as one of the component. To them it appears that such a quantity unnecessarily complicates the expression for the design criterion rather than playing a significant role in designing a discriminatory experiment.

The work of Hsiang and Reilly (1971) is mainly based on the objection that the other criteria fail to allow for prior knowledge, if any, about the values of model parameters. Their personalistic approach utilizes all the information about the degree of belief that one has in rival models and the possible values of their parameters. They have incorporated this information in the form of subjective probability distributions. The procedure laid down by them consists of updating the probabilities of rival models as well as revising the parameter probabilities by Bayes' formula as data come to hand. So far as designing new experiments is concerned, Hsiang and Reilly have proposed the use of Roth's criterion function given by (1.4.3). The implementation of their discrimination method requires storage of discrete values of the parameters of all the models being considered. The method, therefore, is likely to present some computational difficulties if adequate computing facility is not available. Besides, the procedure being entirely based on a particular type of information about the competing models, it may not be possible to use it judiciously if in a given situation the required information is not available. The use of the proposed method is, therefore, restrictive.

Buzzi and Forzatti (1983) belong to a class of researchers in the field of model discrimination who consider the participation of the variances of predictions to be important in designing experiments for discrimination. They have proposed

a criterion which is based on the ratio of two estimates of the variance of residuals under different models, corresponding to an observation due to be realized. According to them, the maximum expected discrimination is attainable at the maximum of this ratio, provided it is rendered greater than unity. To be specific, the new experiment in their sequential procedure may be conducted at a point which maximizes the criterion function

$$\Delta_7(\underset{\sim}{\xi}_{n+1}) = \frac{\sum_{u=1}^{m}\sum_{v=u+1}^{m}\{\eta^{(u)}(\underset{\sim}{\xi}_{n+1},\hat{\underset{\sim}{\theta}}_n^{(u)}) - \eta^{(v)}(\underset{\sim}{\xi}_{n+1},\hat{\underset{\sim}{\theta}}_n^{(v)})\}^2}{(m-1)\{m\hat{\sigma}^2 + \sum_{u=1}^{m}\hat{\sigma}_u^2(\underset{\sim}{\xi}_{n+1})\}}, \qquad (1.4.9)$$

where $\hat{\sigma}^2$ is an estimate of the error variance σ^2 and $\hat{\sigma}_u^2(\underset{\sim}{\xi}_{n+1})$ is an estimate of the variance of prediction made at $\underset{\sim}{\xi}_{n+1}$ through model u, given in Buzzi and Forzatti (1983) .

There may be m models being discriminated but, at any stage the criterion includes only those models which are good from the point of view of adequacy. Since the inadequate models are not allowed to participate in the selection of a new experiment, it is essential to test each model for its consistency with the data available at that stage. To that purpose Buzzi and Forzatti have recommended the use of classical F-test, if an estimate of σ^2 is available and Hartley's F-max or Bartlett's χ^2, if this requirement is not met. Anyhow a model once dropped is likely to be included again if at a later stage it proves to be adequate.

The proposed procedure asks for suspension of designing for discrimination if at any stage Δ_7 could not be maximized above unity by any setting of the input variable(s) in OR. It may, however, be possible to restart the discrimination process if some additional points could be made available for precise estimates of the parameters so as to minimize $\hat{\sigma}_u^2$. This necessitates designing for parameter estimation. According to the authors, the discrimination potency of the procedure can be enhanced by concentrating on the pair of models which are most divergent in the sense used in their work. Besides, a final residual analysis for confirmation of adequacy of the selected model is recommended, in case the F-max or χ^2 statistic have been used for dropping the ill-fitting models during the course of sequential design of experiments.

Buzzi and Forzatti (1983) have pointed out certain advantages which their criterion has over others, especially over that of Box and Hill (1967). For instance, the quantities $\hat{\sigma}_u^2$ being in the denominator of Δ_7 does not affect the criterion value, adversely. The criterion does not give undue importance to the divergence between variances of predictions at the cost of divergence between predictions. Besides, there is no danger of getting such experimental points which will provide a stronger evidence in favour of bad models as these models are kept out of picture while designing for discrimination. Finally, they have claimed that their discrimination procedure is equipped with a desirable stopping rule which some of the other criteria fail to possess.

1.4.2 Review of the Previous Work : Multiresponse Case

The problem of model discrimination is aggravated if, in a given situation, the underlying process happens to be a multiresponse process. In this case, the data from the system come as measurements on two or more responses which must be considered together if one wishes to extract complete information contained in the observations. The situations of this type arise quite naturally in multicomponent processes, such as those involving complex chemical reactions or physical equilibria, etc., and are not uncommon. The models structured for the systems based on these types of processes are usually complex and difficult to handle especially, when these models are nonlinear in parameters. It may also happen that one has to identify an appropriate model from a host of equally plausible models, before it can be used for studying a given system. The investigators in various disciplines frequently face this problem. But, unfortunately, there are only a few techniques available in literature which can be used for discriminating among multivariate models. We now discuss some of the salient features of these techniques.

Roth (1965) is, probably, one of the first few who have considered the multivariate model discrimination problem. In order to design experiments for discriminating among m r-response models he has suggested the use of the weighted function

$$\Delta_8(\underset{\sim}{\xi}_{n+1}) = \prod_{i=1}^{r}\{\sum_{u=1}^{m} P_n^{(u)} \prod_{\substack{v=1 \\ v\neq u}}^{m} |\eta_i^{(u)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(u)}) - \eta_i^{(v)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(v)})|\}, \tag{1.4.10}$$

where $P_n^{(u)}$ is the probability of the truth of model u (also called prior probability) and $\hat{\underset{\sim}{\theta}}_n^{(u)}$ is the m.l.e. of $\underset{\sim}{\theta}^{(u)}$, at the nth stage. The (n+1)th setting of the input variable(s) can be decided by maximizing Δ_8 with respect to $\underset{\sim}{\xi}_{n+1}$. In their sequential procedure, whenever an observation on the r-response random vector is obtained at this designed point the investigator is also supposed to revise the probability of each model through the Bayes' formula. This formula for calculating the posterior probability makes use of the multivariate density of $\underset{\sim}{Y}$ at the (n+1)th run. It may be noted that the proposed procedure is just an extension of his univariate method and, likewise, neglects such errors in divergences as are likely to creep in, if the parameters of the models are not estimated precisely. The design criterion may also give undue importance to models with low probabilities. Besides, the posterior probability itself is not considered to be a reliable tool for making inferences on model adequacy.

One of the several shortcomings that Hsiang and Reilly (1971) have noticed in the Box-Hill procedure is its incapability to handle multiresponse situations. The problem of model discrimination in such situations, in their opinion, can also be tackled through the same Bayesian approach as has been suggested for the univariate models. If the prior information

in the form of model probabilities, the parameter distributions, and the multivariate density of $\underset{\sim}{Y}$ under each model, is available then one could utilize it in designing experiments for the purpose of discrimination. In fact, it is their formula for calculating the posterior probability of a model which combines all this information. It may, however, be noted that this formula is different from that of Box and Hill (1967) but,is used for the same purpose, i.e., for revising the probability of the truth of each model at different stages. The criterion actually proposed for design of experiments is just the product of the values of Δ_3, given in (1.4.3), for each response. Like its univariate analogue this method, too, may present some computational difficulties which may be more serious in a multiresponse situation. The authors have given some suggestions for handling the storage of the data. The proposed procedure does not require error covariance matrix to be known in advance nor does it use linearizations,if the models being discriminated are nonlinear. The discrimination, however, is not possible through the procedure of Hsiang and Reilly if the required information is not available a priori.

Hosten and Froment (1976) have proposed a procedure which is based on the elementary statistical principles. The design criterion suggested therein is a straightforward extension of their own uniresponse criterion with an additional provision of taking the precision of the responses into account. This has been done by incorporating the weights, w_i $(= 1/\sigma_{ii})$, where σ_{ii}

is the variance of response i . According to them, the weighted function

$$\Delta_9(\underset{\sim}{\xi}_{n+1}) = \sum_{i=1}^{r} w_i \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} |\eta_i^{(u)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(u)}) - \eta_i^{(v)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(v)})|, \tag{1.4.11}$$

with $\hat{\underset{\sim}{\theta}}_n^{(u)}$ as the m.l.e. of $\underset{\sim}{\theta}^{(u)}$, provides us with a suitable measure of divergence and can, therefore, be used in designing a discriminatory experiment. As regards analysis of observations obtained from the experiment, Hosten and Froment have used the fact that the statistic

$$\chi^{2(u)} = \sum_{i=1}^{r} \sum_{j=1}^{r} \sigma^{ij} \sum_{k=1}^{n} \{y_{ik} - \eta_i^{(u)}(\underset{\sim}{\xi}_k, \hat{\underset{\sim}{\theta}}_n^{(u)})\}\{y_{jk} - \eta_j^{(v)}(\underset{\sim}{\xi}_k, \hat{\underset{\sim}{\theta}}_n^{(v)})\} \tag{1.4.12}$$

is distributed like Chi-square with $(nr-p_u)$ degrees of freedom. The quantity, σ^{ij}, used in equation (1.4.12) is an (ij)th element of the inverse of the covariance matrix of errors, Σ (assumed to be known in advance). Since the parameter estimation has to be carried out at each stage in order to make $\hat{\underset{\sim}{\theta}}_n^{(u)}$ available for the criterion function at the subsequent stage, the m values of $\chi^{2(u)}$ are generated as by-product and can be used for assessing the adequacy of the rival models. The status of each model is, thus, determined and a model with $\chi^{2(u)}$ greater than the corresponding tabulated value is dropped once for all.

Some of the advantages of this discrimination method lie in the simplicity of the calculations involved and in allowing even the nonlinear models, although the statistic $\chi^{2(u)}$ in that

case would be approximately distributed like Chi-square. Besides, the model finally selected through this method does not suffer from lack of fit. The analysis of experiments, therefore, is not required to be supplemented with additional tests. The discrimination thus achieved has been claimed to result in saving of experimental runs. The procedure of Hosten and Froment has, however, the limitation of being available only for those situations in which Σ is known, a requirement which is rarely met in practice.

While Roth (1965) has neglected the covariance structure of the estimated responses, Hosten and Froment (1976) are opposed to its inclusion in the criterion. Hill and Hunter (1967), on the contrary, have given due importance to the precision and the covariances of the predicted values. They have been able to do so through an extension of the Box-Hill method to the multiresponse case. Likewise, Shannon's concept of entropy has been utilized in a multivariate set-up. In this case, too, it is the upper bound of the expected entropy change which has been exploited for arriving at a criterion function. In order to develop a usable function they have brought in the multivariate equivalents of all those assumptions which have been used by Box and Hill (1967) in the single response case. This has enabled them to arrive at an analogous expression for the discriminatory design criterion, given by

$$\Delta_{10}(\underset{\sim}{\xi}_{n+1}) = \sum_{u=1}^{m-1}\sum_{v=u+1}^{m} P_n^{(u)}P_n^{(v)}[\mathrm{tr}(\Sigma_u\Sigma_v^{-1} + \Sigma_v\Sigma_u^{-1} - 2I_r)$$

$$+ \{\underset{\sim}{\eta}^{(u)}(\underset{\sim}{\xi}_{n+1},\hat{\underset{\sim}{\theta}}_n^{(u)}) - \underset{\sim}{\eta}^{(v)}(\underset{\sim}{\xi}_{n+1},\hat{\underset{\sim}{\theta}}_n^{(v)})\}'(\Sigma_u^{-1} + \Sigma_v^{-1})$$

$$\{\underset{\sim}{\eta}^{(u)}(\underset{\sim}{\xi}_{n+1},\hat{\underset{\sim}{\theta}}_n^{(u)}) - \underset{\sim}{\eta}_n^{(v)}(\underset{\sim}{\xi}_{n+1},\hat{\underset{\sim}{\theta}}_n^{(v)})\}], \qquad (1.4.13)$$

where Σ_u is the precision matrix of model u, Σ_u^{-1} is the inverse of Σ_u, and I_r is an $r \times r$ identity matrix. The criterion obviously makes use of the error structure of the estimated responses through the precision matrix, Σ_u. The decision on the model adequacy is still taken on the basis of the posterior probability, although it now involves the multivariate density of $\underset{\sim}{Y}$ at the (n+1)th run. The method, overall, suffers from certain drawbacks as does its uniresponse analogue. To list a few, the criterion demands a known and constant covariance matrix, Σ; the criterion tends to design points which may be more informative about the inadequate models; the oscillating behaviour of the posterior probability is misleading; and lastly, the procedure tends to favour the simpler model at the cost of the true one. Besides, all those objections which Hsiang and Reilly (1971) have raised against the Box-Hill procedure implicitly hold for its multiresponse version. Whereas many improvements have been devised for the Box-Hill criterion, not much has been done to remove one or more of the above mentioned shortcomings in the discrimination criterion proposed by Hill and Hunter (1967).

The work of Prasad and Rao (1977) is one of the few attempts made to improve upon this procedure. According to them, a more accurate and faster discrimination may be achieved if the expected likelihood is used in lieu of the point likelihood for calculation of posterior probabilities of models. This not only results into a sharp discrimination but also selects better points through the criterion function, Δ_{10}. The development of such an alternative expression for the model likelihood is based on the simple idea that likelihood, being a function of the parameters, is a random variable and can, therefore, be subjected to the expectation operator. So far as design of experiments is concerned, Prasad and Rao have recommended the use of Hill-Hunter criterion only. This means that the only difference between the two methods lies in the formulae for posterior probability : Hill and Hunter have used the point likelihood, while Prasad and Rao have recommended the use of the expected likelihood. The performance of the two alternative formulae has been compared by Prasad and Rao (1977) through an application to a practical example involving eleven bivariate models. With the same prior information and initial conditions, the use of the expected likelihood has been found to be more decisive so far as the rejection of the ill-fitting models is concerned. It has also been shown through this example that the convergence towards complete discrimination is faster with the expected likelihood than it is with the point likelihood. The authors have, however, pointed out that the

superiority of their method depends on the models as well as on the type of data. The modification introduced by them seems to have improved the rate of discrimination of the multiresponse version of the Box-Hill procedure, but has not removed the basic drawbacks.

Buzzi and Forzatti (1983) have claimed to remove certain limitations and anomalies of the Box-Hill procedure for univariate models through an entirely different approach. The work of Buzzi et al. (1984) is a similar attempt in this direction for the multiresponse case. In fact, they have subjected the Hill-Hunter method to a similar criticism and developed a method for multiresponse model discrimination which happens to be a generalization of the univariate method of Buzzi and Forzatti (1983). The criterion is based on the minimization of the likelihood function of the divergences under any pair of models. As this is equivalent to maximizing the exponent involved, a convenient function for the purpose of designing discrimminatory experiments is considered to be

$$\Delta_{11}(\underset{\sim}{\xi}_{n+1}) = [\underset{\sim}{\eta}^{(u)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(u)}) - \underset{\sim}{\eta}^{(v)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(v)})]' \Sigma_{u,v}^{-1}$$

$$[\underset{\sim}{\eta}^{(u)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(u)}) - \underset{\sim}{\eta}^{(v)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}_n^{(v)})], \qquad (1.4.14)$$

where $\Sigma_{u,v} = \Sigma + \Sigma_u + \Sigma_v$. In a problem of discriminating among m r-variate models the next experiment, in the sequential procedure of Buzzi et al., is recommended to be conducted at a point which maximizes Δ_{11} over any pair of models, u,v, provided

$\Delta_{11}(\underset{\sim}{\xi}_{n+1})$ is greater than the number of responses involved. The design for discrimination is continued till this inequality is satisfied. The emphasis is, thereafter, shifted to design of some experiments for precise estimation of model parameters. This may make the restart of the discrimination process possible. While the adequacy of all the models participating in discrimination, at each stage, is recommended to be checked by χ^2-test, the procedure is finally wound up by carrying out residual analysis of the selected model. It may be noted that the implementation of this technique requires multivariate normal distribution of errors with known covariance matrix, Σ. Besides, the models being discriminated are required to be linear or linearized in the parameter space. The procedure shares all the advantages and disadvantages of its univariate analogue.

CHAPTER 2

METHODOLOGY FOR MODEL DISCRIMINATION: DESIGN AND ANALYSIS

2.1 AN APPROACH TO THE PROBLEM OF MODEL DISCRIMINATION

A mechanistic model is an abstract formulation aimed at simulating the mechanism of a real world phenomenon. The basic step involved in building such a model for a system is, therefore, to establish the true nature of its process. To that purpose, it may be borne in mind that so far as observations are concerned, randomness is somehow imbued into the process through one or the other sources, thus, rendering the process probabilistic in nature. Accordingly, if one endeavours to build a model for such a system one must consider the random error, in addition to the parameters (associated with the mechanism of the process) and a set of exogeneous variables (characterizing the input into the system). Mathematically, an r-response ($r \geq 1$) system can, therefore, be reasonably represented by a set of r algebraic equations

$$y_i = \eta_i(\underset{\sim}{\xi}, \underset{\sim}{\theta}) + \varepsilon_i \ , \quad i = 1,2,\ldots,r, \qquad (2.1.1)$$

where $\eta_i(\underset{\sim}{\xi}, \underset{\sim}{\theta})$ stands for the true value and y_i represents the observed value of the ith response, $\underset{\sim}{\xi}$ is a $q \times 1$ vector of input

variables, $\underset{\sim}{\theta}$, a p x1 vector, is representative of the physical parameters of the system, and ε_i, assumed to be additive to the true response, accounts for the variability in observations.

Model (2.1.1), to be denoted as $M^{(o)}$, is fundamentally a characterization of some random variable(s), $\underset{\sim}{Y}$, of interest, in terms of the parameters $\underset{\sim}{\theta}$ and independent variable(s) $\underset{\sim}{\xi}$, while the observations arising from the system as a result of certain inputs can be presumed to be realizations of $\underset{\sim}{Y}$. Thus, whereas it is reasonable to assume a probability distribution associated with this model, it is as well justifiable to assume the existence of a population consisting of observations on the characteristic(s), $\underset{\sim}{Y}$, being studied in a given situation. Let $df^{(o)}$ be the probability distribution of $\underset{\sim}{Y}$ corresponding to model $M^{(o)}$ and $\prod^{(o)}$ be the population which is supposed to have been generated by the underlying process, $\Omega^{(o)}$.

Unfortunately, a true mathematical model, such as (2.1.1), fully accounting for the underlying mechanism of the process is a hypothetical concept and can seldom be realized in non-trivial situations. All that one ought to look for is, therefore, a set of mathematical equation(s) which can behave considerably akin to the process, if not exactly. In other words, a simpler but closely approximating model which could describe the outstanding features of (2.1.1) is sought for. We consider a situation in which several models, m (say), have been hypothesized to possess this qualification. These models are claimed to be capable of

describing the mechanism of the given system equally well and are, therefore, taken as rivals,so far as the problem of model identification is concerned. Each potential model in this case can be entertained as having possibly generated the data. In other words, all the rival models can be presumed to offer a satisfactory explanation for the observations being taken on the random variable(s), $\underset{\sim}{Y}$. It may thus be argued that $\underset{\sim}{Y}$ can, alternatively, be characterized by model, $M^{(u)}$ (say), in terms of the parameters $\underset{\sim}{\theta}^{(u)}$ and the input variables $\underset{\sim}{\xi}$. Thus, if we prefer to represent the given system through $M^{(u)}$ then we will have

$$y_i = \eta_i^{(u)} (\underset{\sim}{\xi}, \underset{\sim}{\theta}^{(u)}) + \varepsilon_i^{(u)} , \qquad i = 1,2,\ldots,r. \qquad (2.1.2)$$

Let $df^{(u)}$ be the probability distribution of $\underset{\sim}{Y}$ associated with this model.

Now, considered independently, each of these models, model u for instance, must be a true model for some process, $\cap^{(u)}$ (say), which may be visualized as generating a population $\prod^{(u)}$ (say). Accordingly, the process $\cap^{(u)}$ can be represented by model u through r algebraic equations

$$y_i^{(u)} = \eta_i^{(u)} (\underset{\sim}{\xi}, \underset{\sim}{\theta}^{(u)}) + \varepsilon_i^{(u)} , \quad i = 1,2,\ldots,r \ (r \geq 1), \quad u = 1,2,\ldots,m \ (m \geq 2), \qquad (2.1.3)$$

where $y_i^{(u)}$ stands for the observed value and $\eta_i^{(u)} (\underset{\sim}{\xi}, \underset{\sim}{\theta}^{(u)})$ represents the true value of the ith response, from $\cap^{(u)}$,

depending on $\underset{\sim}{\xi}$ and a set of p_u parameters $\underset{\sim}{\theta}^{(u)}$, while $\varepsilon_i^{(u)}$ represents the random factor operating on the process $\cap^{(u)}$. It may, however, be noted that the values $y_i^{(u)}$ in (2.1.3) cannot be actually observed because the process $\cap^{(u)}$ is not known; it has been assumed to exist hypothetically. Whatsoever, by virtue of the argument used, the values simulated through (2.1.2) may be treated as belonging to the population $\prod^{(u)}$.

Now, the main purpose of building a mechanistic model is to be able to replace the given system by certain mathematical equation(s) which can simulate such observations as if these have been generated by the underlying process, $\cap^{(o)}$, itself. This suggests that a model is capable of best representing the given process if the population simulated by it is considerably close (in some statistical sense) to the population associated with the process $\cap^{(o)}$. Therefore, the discrepency of the model, $M^{(u)}$ (say), if any, can be seen through the dissimilarity of $\prod^{(u)}$ from $\prod^{(o)}$. In other words, in order to check the credibility of model u we must look for the closeness between these two populations. Since the statistical populations, i.e., the type of populations being considered here, can be well characterized by their frequency distributions, the akinness between $\prod^{(o)}$ and $\prod^{(u)}$ can be appropriately assessed through the identicalness between their probability distributions.

In a given situation the aim may also be to acquire a discriminatory observation, i.e., to append the available sample

of n observations with another, (n+1)th, which could add to the sharpness in discrimination between models $M^{(u)}$ and $M^{(v)}$, for instance. It may similarly be argued that in this case it is the divergence between the probability distributions $df_{n+1}^{(u)}$ and $df_{n+1}^{(v)}$ (with reference to future), under models u and v, of some random variable(s), $\underset{\sim}{Y}_{n+1}$, which may be utilized. In case of m models there will be m such probability distributions. Any divergence among these probability distributions is expected to result into more clear a distinction among the underlying models.

In this work, for the purpose of discrimination and designing experiments conducive to discrimination we, therefore, plan to employ a distance function through which we could evaluate the extent of identicalness between two probability distributions.

2.2 SEARCH FOR AN APPROPRIATE FUNCTION

Let $(\mathcal{X}, \mathcal{B}, \nu)$ be a measure space and $\mathcal{P}$ be the set of all probability measures on $\mathcal{B}$ which are absolutely continuous with respect to ν. Consider two such probability measures, $p_1, p_2 \in \mathcal{P}$ and let f_1, f_2 be the corresponding probability density functions such that $f_1 = dp_1/d\nu$ and $f_2 = dp_2/d\nu$. Then one of the measures of affinity between the probability distributions corresponding to f_1 and f_2 is defined as [Hellinger (1909)]

$$h(f_1, f_2) = \int (f_1 f_2)^{1/2} \, d\nu \, . \qquad (2.2.1)$$

But, symmetry is the only property which is satisfied by the function, h. Therefore, it does not qualify to be a metric. Since specific to our requirement we need a distance function, h is not an appropriate function for discrimination. However, the function K, defined as

$$K(f_1,f_2) = [1 - h(f_1,f_2)]^{1/2} \quad , \qquad (2.2.2)$$

can be verified to be a distance function and is, therefore, suitable for the present purpose.

2.2.1 Metric Properties of the Function, K

We now prove the metric properties of K .

(i) $K(f_1,f_2) \geq 0$.

Clearly, $0 \leq f_1,f_2 \leq 1$ a.e. ν .

This leads to the inequality,

$$0 \leq h(f_1,f_2) \leq 1 \qquad \text{a.e. } \nu$$

and hence to the inequality,

$$K(f_1,f_2) \geq 0 \ .$$

In fact, we have

$$0 \leq K(f_1,f_2) \leq 1.$$

(ii) $K(f_1,f_2) = 0$ if and only if $f_1 = f_2$ a.e. ν .

Let $f_1 = f_2$. Then

$$h(f_1,f_2) = \int f_1 \, d\nu$$

$$= 1.$$

Therefore,

$$K(f_1,f_2) = \{1 - h(f_1,f_2)\}^{1/2}$$

$$= 0.$$

Conversely, suppose $K(f_1,f_2) = 0$.

This implies $\{1-h(f_1,f_2)\}^{1/2} = 0$

and, in turn, we have

$$h(f_1,f_2) = 1$$

which is possible only if $f_1 = f_2$ a.e. ν .

(iii) $K(f_1,f_2) = K(f_2,f_1)$. a.e. ν .

The symmetry of K follows from that of h.

(iv) Consider three probability measures $p_1,p_2,p_3 \in \mathcal{P}$ which are absolutely continuous with respect to ν. Let f_1,f_2,f_3 be the corresponding probability density functions such that $f_i = dp_i/d\nu$, $i = 1,2,3$. Then,

$$K(f_1,f_3) \leq K(f_1,f_2) + K(f_2,f_3).$$

By definition,

$$K(f_1,f_3) = \{1 - h(f_1,f_3)\}^{1/2}$$

$$= \frac{1}{\sqrt{2}} \{2 - 2h(f_1,f_3)\}^{1/2}$$

$$= \frac{1}{\sqrt{2}} \{\int f_1 d\nu + \int f_3 d\nu - 2 \int (f_1 f_3)^{1/2} d\nu\}^{1/2}$$

$$= \frac{1}{\sqrt{2}} \{\int (f_1^{1/2} - f_3^{1/2})^2 \, d\nu - 2 \int (f_1 f_3)^{1/2} d\nu\}^{1/2}$$

$$= \frac{1}{\sqrt{2}} \}\int (f_1^{1/2} - f_3^{1/2})^2 \, d\nu\}^{1/2}$$

$$= \frac{1}{\sqrt{2}} [\int \{(f_1^{1/2} - f_2^{1/2}) + (f_2^{1/2} - f_3^{1/2})\}^2 \, d\nu]^{1/2}.$$

Using Minkowski's inequality on the right hand side we get

$$K(f_1,f_3) \leq \frac{1}{\sqrt{2}} [\{\int (f_1^{1/2} - f_2^{1/2})^2 d\nu\}^{1/2} + \{\int (f_2^{1/2} - f_3^{1/2})^2 d\nu\}^{1/2}]$$

$$= \{1-h(f_1,f_2)\}^{1/2} + \{1-h(f_2,f_3)\}^{1/2},$$

i.e.,

$$K(f_1,f_3) \leq K(f_1,f_2) + K(f_2,f_3).$$

The properties (i), (ii), (iii), and (iv) show that K is a distance function. With this the search for an appropriate function is over.

2.2.2 Utilizing the Distance Function, K, in Model Discrimination

We now recall the hierarchical relation among the model $M^{(u)}$, the population $\prod^{(u)}$, and the corresponding probability distributions $df^{(u)}$, $u = 0,1,2,\ldots,m$. Keeping in view this relation and the fact that K is a suitable measure of identicalness between two probability distributions, as established earlier, it can be appreciated that K may be utilized in measuring the dissimilarity between $\prod^{(o)}$ and any one of the m populations, $\prod^{(1)}, \prod^{(2)}, \ldots, \prod^{(m)}$. This function may, therefore, be utilized in ariving at a scale for discriminating among the rival models.

According to the concept used in this work, discrimination is, therefore, proposed to be done by means of a sample estimate of the distance between $\prod^{(o)}$ and $\prod^{(u)}$, u = 1,2,...,m. We will denote this estimate by $K_n^{(u)}$, when calculated through n values of $\underset{\sim}{Y}$. Besides, on the basis of the properties possessed by the function K, it may also be employed in diverging the probability distributions of the random variable(s), $\underset{\sim}{Y}_{n+1}$, under the models being discriminated and hence in designing discriminating experiments. The distance between $df_{n+1}^{(u)}$ and $df_{n+1}^{(v)}$, in this case will be denoted as $K_{u,v}(\underset{\sim}{\xi}_{n+1})$, $\underset{\sim}{\xi}_{n+1}$ being the (n+1)th setting of the experiment yet to be conducted.

2.3 MODEL DISCRIMINATION : NONSEQUENTIAL

In practice the investigator might have to choose a model from amongst a set of proposed models on the basis of a fixed number of observations only. Situations of this type arise when the experimental apparatus has already been dismanteled or the cost of conducting another run of the experiment is prohibitively high. This rules out the possibility of continuing experimentation beyond n runs. Thus the decision on the choice of best model(s) has to be taken on the basis of whatever is available at hand. In this case whatever discrimination one could achieve through the available data would be worth it, for one would be working under such a constraint that furthering of discrimination through the addition

of more informative experiments is not possible. Therefore, even if one ends up with a reasonably small number of closest models, the rest of the task can be accomplished through other considerations, such as cost and convenience etc. in using a particular model from the group of better ones. In a given situation, if there are m models being discriminated, it may be possible to select a subset of m_1 ($< m$) models by comparing m values of $K_n^{(u)}$, $u = 1,2,\ldots,m$, while the selection of a single model may not be possible on the basis of a fixed number of observations. However, with m_1 as a small number (say 2 or 3) the final selection may be made through some of the practical considerations relevant to the given situation.

2.4 MODEL DISCRIMINATION : SEQUENTIAL

Suppose that one starts with a certain number of observations, n (say) but, these do not prove to be carrying enough information so as to enable one to take a decision on the best model. Now, if unlike the previous case, there is a scope of additional experimentation, it may be possible to achieve a better discrimination, often arriving at a single model. A few more observations appended, sequentially, to the sample of n are expected to do the job. However, in this case, in order to achieve this objective the investigator needs two criteria : a discrimination criterion, for making an assessment of the discrimination achieved at every stage so that only the minimum

number of experiments are conducted, and a design criterion which would suggest the setting of the input variable(s) at which an experiment, beyond nth stage, should be conducted so as to fetch more information for discrimination.

2.4.1 Discrimination Criterion : An Index

So far as the decision on the relative adequacy of models at the nth stage is concerned, we propose to use the statistic, $D_n^{(u)}$, defined as

$$D_n^{(u)} = \frac{D_{n-1}^{(u)} K_n^{(u)}}{\sum_{v=1}^{m} D_{n-1}^{(v)} K_n^{(v)}} , \quad u = 1,2,\ldots,m, \qquad (2.4.1)$$

where $D_{n-1}^{(u)}$ is the value of the statistic for model u at the previous, (n-1)th, stage and $K_n^{(u)}$ measures the discrepency of this model in explaining the mechanism of the given process at the current, nth, stage. The proposed statistic will, hereafter, be referred to as Discrimination Index. Since in model discrimination problems we start with the assumption that all the proposed models are equally plausible, it is reasonable to assume, initially, that these models present equally close approximations to the process model, $M^{(o)}$. This amounts to assuming that $D_{n-1}^{(u)} = 1/m$, $u = 1,2,\ldots,m$. The discrimination index thus defined makes use of the discrimination realized at a given stage as well as the discrimination already achieved through the previous stages. Besides, the value of the index will lie between 0 and 1, with both values inclusive.

2.4.2 Design Criterion : A Weighted Function

If at the nth stage, the m values of the discrimination index do not show enough evidence in favour of a single model and further experimentation is possible, the investigator must conduct a few more experiments specific to the purpose of discrimination. This necessitates the use of a design criterion. The criterion that we propose for designing discriminative experiments is based on the distance function K, as defined by (2.2.2). In fact, for deciding on the (n+1)th setting, $\underset{\sim}{\xi}_{n+1}$, we consider the random variable(s) $\underset{\sim}{Y}_{n+1}$, on which a discriminatory observation is yet to be realized. Thus, in a problem of discriminating between two models, for example, one could obtain a design point by maximizing the function

$$K_{1,2}(\underset{\sim}{\xi}_{n+1}) = [1 - h(f_{n+1}^{(1)}, f_{n+1}^{(2)})]^{1/2} \qquad (2.4.2)$$

with respect to $\underset{\sim}{\xi}_{n+1}$, where $f_{n+1}^{(1)}$, and $f_{n+1}^{(2)}$ are the p.d.f.'s of $\underset{\sim}{Y}_{n+1}$ under models 1 and 2, respectively. But, if there are m models being discriminated there will be $\binom{m}{2}$ pairwise distances,

$$K_{u,v}(\underset{\sim}{\xi}_{n+1}) = [1 - h(f_{n+1}^{(u)}, f_{n+1}^{(v)})]^{1/2} , \qquad (2.4.3)$$

with $f_{n+1}^{(u)}$ and $f_{n+1}^{(v)}$ as the p.d.f.'s of Y_{n+1} under models u and v, respectively. We will, alternatively, use $K_{u,v;n+1}$ to denote the distance such as given by (2.4.3). According to the approach we plan to use here, the design point yielding the most informative experiment for bringing the best model to light would be the one which simultaneously maximized all the distances $K_{u,v;n+1}$. But, unfortunately, a point which maximizes one distance may not

maximize another. There is, thus, the necessity of choosing appropriate weights which represent the importance of various distances in designing new experiments and then combining all the distances into a weighted function. The weights, in fact, determine the role of the distances, involved, im design of experiments.

The Weights : The weights, which we plan to use, are based on the strategy that the models which are closer should receive more attention than the ones which are comparatively farther. This way a pair with the closest models in it would be attached with the highest weight while the one comprising of farthest models is given the least importance. One such set of weights may be proposed as

$$w_{u,v;n} = \frac{D_n^{(u)}}{\tau_1 D_n^{(v)}} \quad \text{if } D_n^{(u)} \leq D_n^{(v)},$$

$$= \frac{D_n^{(v)}}{\tau_2 D_n^{(u)}} \quad \text{otherwise,} \qquad (2.4.4)$$

where τ_1 and τ_2 are the corresponding normalizing constants.

The Criterion function : So far as the criterion function is concerned, we prefer the use of the weighted average of the $\binom{m}{2}$ pairwise distances, the weights being given by equations(2.4.4). To be more specific, in a problem of designing experiments for discriminating among m models we propose to choose the next design point, $\underset{\sim}{\xi}_{n+1}$, at the maximum of

$$\phi(\underset{\sim}{\xi}_{n+1}) = \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} w_{u,v;n} K_{u,v}(\underset{\sim}{\xi}_{n+1}), \qquad (2.4.5)$$

where $K_{u,v}(\underset{\sim}{\xi}_{n+1})$ is given by (2.4.3). It may be noted that the search for the maximum is always made over the operability region. An experiment conducted at such a setting of the input variable(s) is expected to yield the most informative (from discrimination point of view) observation on the response(s) of the system.

2.5 DEVELOPMENT OF THE DISCRIMINATION CRITERION

2.5.1 Some Assumptions : It has already been proposed that the discrimination among models can be done through the population, $\prod^{(o)}$, associated with the process, $\eta^{(o)}$ on the one hand and the populations, $\prod^{(1)}, \prod^{(2)}, \ldots, \prod^{(m)}$, supposed to have been simulated through the competing models on the other. Besides, it has also been argued that we have reasons to associate a probability distribution with each of these populations. In fact, we designated their associated distributions as $df^{(o)}$; $df^{(1)}, df^{(2)}, \ldots, df^{(m)}$. In this section, we proceed further to specify the particular forms to these distributions and thus give a concrete shape to the laid down strategy for discrimination among the underlying models.

It may be recalled that whereas the population, $\prod^{(o)}$, consists of the values of certain random variable(s) observed from the system, the elements of the population, $\prod^{(u)}$, are the ones simulated through model, $M^{(u)}$, $u = 1,2,\ldots,m$. Theoretically speaking, these populations can be made infinitely large by continuing the experimentation and by carrying out the simulation

through the models, correspondingly. We assume that these populations have normal distributions; an assumption which can be justified in this case. Thus, in the univariate case, whereas we let $\prod^{(o)}$ have the distribution, $N_1(\alpha^{(o)}, \lambda^{(o)})$, the distribution of $\prod^{(u)}$ may be specified as $N_1(\alpha^{(u)}, \lambda^{(u)})$, $u = 1,2,\ldots,m$. Similarly, when the underlying populations are multivariate, we assume that $\prod^{(o)}$ has an r-variate normal distribution : $N_r(\underset{\sim}{\alpha}^{(o)}, \Lambda^{(o)})$, while the distribution $N_r(\underset{\sim}{\alpha}^{(u)}, \Lambda^{(u)})$ is appropriate for the population $\prod^{(u)}$, $u = 1,2,\ldots,m$. The development so far bring to light the important point that the proposed discrimination criterion is to be based on the identicalness of normal distributions : $N_1(\alpha^{(o)}, \lambda^{(o)})$ and $N_1(\alpha^{(u)}, \lambda^{(u)})$ or $N_r(\underset{\sim}{\alpha}^{(o)}, \Lambda^{(o)})$ and $N_r(\underset{\sim}{\alpha}^{(u)}, \Lambda^{(u)})$, $u = 1,2,\ldots,m$, according as the system is uniresponse or multiresponse. We, therefore, derive below the explicit expressions for the distance function, K, defined for two p.d.f.'s under the Gaussian set-up, for the univariate and multivariate cases, respectively.

2.5.2 Evaluating the Akinness Between Two Normal Populations

Case 1 : Univariate Populations ;

Result 1 : Let $\prod_1$ and $\prod_2$ be two univariate populations with normal probability distributions : $N_1(\alpha_1, \lambda_1)$ and $N_1(\alpha_2, \lambda_2)$, respectively. Then the dissimilarity between these populations is given by

$$K(\Pi_1, \Pi_2) = [1 - (\frac{\lambda_1\lambda_2}{\lambda^4})^{1/4} \exp\{-\frac{1}{8}(\frac{\alpha_1-\alpha_2}{\lambda})^2\}]^{1/2}, \tag{2.5.1}$$

where

$$\lambda = (\frac{\lambda_1+\lambda_2}{2})^{1/2}. \tag{2.5.2}$$

Proof : Let f_1 and f_2 denote the p.d.f.'s of the random variable Y corresponding to the population, Π_1 and Π_2, respectively, so that

$$f_i(y) = [2\pi\lambda_i]^{-1/2} \exp[-\frac{1}{2}\frac{(y-\alpha_i)^2}{\lambda_i}], \quad -\infty < y < \infty, \quad i = 1,2. \tag{2.5.3}$$

We first evaluate the function, h, given by

$$h(f_1,f_2) = \int_{R^1} [f_1(y)f_2(y)]^{1/2}\, dy. \tag{2.5.4}$$

Using (2.5.3) in (2.5.4), we have

$$h(f_1,f_2) = [\lambda_1\lambda_2]^{-1/4}(2\pi)^{-1/2} \int_{R^1} \exp[-\frac{1}{4}\{\frac{(y-\alpha_1)^2}{\lambda_1} + \frac{(y-\alpha_2)^2}{\lambda_2}\}]dy. \tag{2.5.5}$$

By relation (A.1.11) of Lemma A.1 (Appendix A), the expression in the exponent of the integrand can be written as

$$\frac{(y-\alpha_1)^2}{\lambda_1} + \frac{(y-\alpha_2)^2}{\lambda_2} = \frac{(y-\alpha^*)^2}{\lambda^*} + \frac{(\alpha_1-\alpha_2)^2}{\lambda_1+\lambda_2},$$

where $\alpha^* = \dfrac{\lambda_2\alpha_1 + \lambda_1\alpha_2}{\lambda_1+\lambda_2}$ and

$$\lambda^* = \left(\frac{\lambda_1 \lambda_2}{\lambda_1+\lambda_2}\right). \tag{2.5.6}$$

Using these relations in (2.5.5), we obtain

$$h(f_1,f_2) = [\lambda_1\lambda_2 (2\lambda^*)^{-2}]^{-1/4} \exp\left[-\frac{1}{4}\left\{\frac{(\alpha_1-\alpha_2)^2}{\lambda_1+\lambda_2}\right\}\right] \times$$

$$[2\pi(2\lambda^*)]^{-1/2} \int_{R^1} \exp\left[-\frac{1}{2}\left\{\frac{(y-\alpha)^2}{2\lambda^*}\right\}\right] dy$$

$$= [\lambda_1\lambda_2 (2\lambda^*)^{-2}]^{-1/4} \exp\left[-\frac{1}{4}\left\{\frac{(\alpha_1-\alpha_2)^2}{\lambda_1+\lambda_2}\right\}\right]. \tag{2.5.7}$$

Substituting for λ^* from equation (2.5.6) we get

$$h(f_1,f_2) = \left[\frac{(\lambda_1+\lambda_2)^2}{4\lambda_1\lambda_2}\right]^{-1/4} \exp\left[-\frac{1}{4}\left\{\frac{(\alpha_1-\alpha_2)^2}{\lambda_1+\lambda_2}\right\}\right]. \tag{2.5.8}$$

Now, if we let $\lambda = \left(\frac{\lambda_1+\lambda_2}{2}\right)^{1/2}$, then h can be written in a nice form like

$$h(f_1,f_2) = \left[\frac{\lambda_1\lambda_2}{\lambda^4}\right]^{1/4} \exp\left[-\frac{1}{8}\left(\frac{\alpha_1-\alpha_2}{\lambda}\right)^2\right]$$

so that, for the given situation, the distance function, K, defined in (2.2.2), can be expressed in the form

$$K(\Pi_1, \Pi_2) = \left[1 - \left(\frac{\lambda_1\lambda_2}{\lambda^4}\right)^{1/4} \exp\left\{-\frac{1}{8}\left(\frac{\alpha_1-\alpha_2}{\lambda}\right)^2\right\}\right]^{1/2}.$$

With a specific interest we also write an alternative

form of $h(f_1,f_2)$; namely,

$$h(f_1,f_2) = \left[\frac{\lambda^4}{\lambda_1\lambda_2}\right]^{-1/4} \exp\left[-\frac{1}{8}\left(\frac{\alpha_1 - \alpha_2}{\lambda}\right)^2\right], \qquad (2.5.9)$$

so as to evaluate the function,

$$L(\Pi_1, \Pi_2) = -\log_e h(f_1,f_2).$$

In fact, on using equation (2.5.9) we get

$$L(\Pi_1, \Pi_2) = \frac{1}{8}D_1(\alpha_1,\alpha_2) + \frac{1}{4}D_2(\lambda_1,\lambda_2), \qquad (2.5.10)$$

where

$$D_1(\alpha_1,\alpha_2) = \left(\frac{\lambda_1 + \lambda_2}{2}\right)^{-1} (\alpha_1 - \alpha_2)^2 \qquad (2.5.11)$$

and

$$D_2(\lambda_1,\lambda_2) = 2\log_e\left(\frac{\lambda_1 + \lambda_2}{2}\right) - \log_e \lambda_1 - \log_e \lambda_2 . \qquad (2.5.12)$$

<u>Case</u> 2 : <u>Multivariate Populations</u>

<u>Result 2</u> : Let Π_1 and Π_2 be two r-variate normal populations with non singular r-dimensional probability distributions : $N_r(\underset{\sim}{\alpha}_1, \Lambda_1)$ and $N_r(\underset{\sim}{\alpha}_2, \Lambda_2)$. Then the dissimilarity between Π_1 and Π_2 is given by

$$K(\Pi_1, \Pi_2) =$$

$$\left[1 - \left(\frac{|\Lambda_1\Lambda_2|}{|\Lambda|^2}\right)^{1/4} \exp\left\{-\frac{1}{8}(\underset{\sim}{\alpha}_1 - \underset{\sim}{\alpha}_2)' \Lambda^{-1}(\alpha_1 - \alpha_2)\right\}\right]^{1/2}, \qquad (2.5.13)$$

where

$$\Lambda = (\Lambda_1 + \Lambda_2)/2. \qquad (2.5.14)$$

Proof : Since the probability distribution of Π_i is r-variate normal : $N_r(\underset{\sim}{\alpha}_i, \Lambda_i)$, i = 1,2, the probability density functions, f_1 and f_2, (say) of a random vector $\underset{\sim}{Y}$ (say) are given by

$$f_i(\underset{\sim}{y}) = [(2\pi)^r \ |\Lambda_i|]^{-1/2} \exp\{-\tfrac{1}{2} (\underset{\sim}{y}-\underset{\sim}{\alpha}_i)' \Lambda_i^{-1} (\underset{\sim}{y}-\underset{\sim}{\alpha}_i)], \ i=1,2, \tag{2.5.15}$$

where Λ_i is r x r positive definite symmetric covariance matrix, $|\Lambda_i|$ denotes the determinant, Λ_i^{-1} stands for its inverse, and $\underset{\sim}{y}$, $\underset{\sim}{\alpha}_i$ are r-vectors. By definition, the distance function, K, is given by

$$K(\Pi_1,\Pi_2) = [1 - \int_{R^r} \{f_1(\underset{\sim}{y}) \ f_2(\underset{\sim}{y})\}^{1/2} \ d\underset{\sim}{y}]^{1/2}. \tag{2.5.17}$$

Substituting for $f_1(y)$ and $f_2(y)$ from equations (2.5.15) in the function

$$h(f_1,f_2) = \int_{R^r} \{f_1(\underset{\sim}{y}) f_2(\underset{\sim}{y})\}^{1/2} \ d\underset{\sim}{y} \ ,$$

we have

$$h(f_1,f_2) = [(2\pi)^{2r} \ |\Lambda_1| \ |\Lambda_2|]^{-1/4}$$

$$\int_{R^r} \exp[-\tfrac{1}{4}\{(\underset{\sim}{y}-\underset{\sim}{\alpha}_1)' \Lambda_1^{-1}(\underset{\sim}{y}-\underset{\sim}{\alpha}_1)+(\underset{\sim}{y}-\underset{\sim}{\alpha}_2)' \Lambda_2^{-1}(\underset{\sim}{y}-\underset{\sim}{\alpha}_2)\}] \ d\underset{\sim}{y}. \tag{2.5.17}$$

Using equations (A.1.1), (A.1.2), and (A.1.3) of Lemma A.1 (Appendix A), we can write

$$(\underset{\sim}{y}-\underset{\sim}{\alpha}_1)' \Lambda_1^{-1}(\underset{\sim}{y}-\underset{\sim}{\alpha}_1) + (\underset{\sim}{y}-\underset{\sim}{\alpha}_2)' \Lambda_2^{-1}(\underset{\sim}{y}-\underset{\sim}{\alpha}_2) =$$

$$(\underset{\sim}{y}-\underset{\sim}{\alpha}^*)' \Lambda^* (\underset{\sim}{y}-\underset{\sim}{\alpha}^*) + (\underset{\sim}{\alpha}_1-\underset{\sim}{\alpha}_2)' (\underset{\sim}{\Lambda}_1+\underset{\sim}{\Lambda}_2)^{-1} (\underset{\sim}{\alpha}_1-\underset{\sim}{\alpha}_2),$$

where $\underset{\sim}{\alpha}^* = (\Lambda_1+\Lambda_2)^{-1}\ (\Lambda_2\underset{\sim}{\alpha}_1 + \Lambda_1\underset{\sim}{\alpha}_2)$

and $\Lambda^* = \Lambda_1^{-1}\ (\Lambda_1 + \Lambda_2)\ \Lambda_2^{-1}.$

These relations when used in (2.5.17) give

$$h(f_1,f_2) = \left[\frac{4|\Lambda_1\Lambda_2|}{|\Lambda_1+\Lambda_2|^2}\right]^{1/4} \exp\left[-\tfrac{1}{4}(\underset{\sim}{\alpha}_1-\underset{\sim}{\alpha}_2)'\ (\Lambda_1+\Lambda_2)^{-1}\ (\underset{\sim}{\alpha}_1-\underset{\sim}{\alpha}_2)\right].$$

And, if we,let $\Lambda = (\Lambda_1 + \Lambda_2)/2$,i.e., pool the covariance matrices, then h can be written as

$$h(f_1,f_2) = \left[\frac{|\Lambda|^2}{|\Lambda_1\Lambda_2|}\right]^{-1/4}\exp\left[-\tfrac{1}{8}\ (\underset{\sim}{\alpha}_1-\underset{\sim}{\alpha}_2)'\ \Lambda^{-1}\ (\underset{\sim}{\alpha}_1-\underset{\sim}{\alpha}_2)\right]. \quad (2.5.18)$$

Accordingly from (2.2.2) we obtain the dissimilarity between Π_1 and Π_2 in the form

$$K(\Pi_1,\Pi_2) = \left[1-\left(\frac{|\Lambda_1\Lambda_2|}{|\Lambda|^2}\right)^{1/4} \exp\{-\tfrac{1}{8}(\underset{\sim}{\alpha}_1-\underset{\sim}{\alpha}_2)'\ \Lambda^{-1}(\underset{\sim}{\alpha}_1-\underset{\sim}{\alpha}_2)\}\right]^{1/2}, \quad (2.5.19)$$

where

$$\Lambda = (\Lambda_1+\Lambda_2)/2. \quad (2.5.20)$$

In addition, we find that the function L, defined as

$$L(\Pi_1,\Pi_2) = -\log_e h(f_1,f_2),$$

becomes

$$L(\Pi_1,\Pi_2) = \tfrac{1}{8}D_1(\underset{\sim}{\alpha}_1,\underset{\sim}{\alpha}_2) + \tfrac{1}{4}D_2(\Lambda_1,\Lambda_2), \quad (2.5.21)$$

where

$$D_1(\underset{\sim}{\alpha}_1,\underset{\sim}{\alpha}_2) = \left[(\underset{\sim}{\alpha}_1-\underset{\sim}{\alpha}_2)'\ \left(\frac{\Lambda_1+\Lambda_2}{2}\right)^{-1}\ (\underset{\sim}{\alpha}_1-\underset{\sim}{\alpha}_2)\right] \quad (2.5.22)$$

and $D(\Lambda_1, \Lambda_2) = 2 \log_e(|\frac{\Lambda_1+\Lambda_2}{2}|) - \log_e|\Lambda_1| - \log|\Lambda_2|$. (2.5.23)

2.5.3 An Important Observation

Although our main interest in both the cases: univariate and multivariate, is in the function K, the function L, given in equations (2.5.10) and (2.5.21), has been worked out in order to bring out an important point. Consider again the populations Π_1, and Π_2. Since these populations have been assumed to be r-variate normal statistical populations ($r \geq 1$), each can be described by an r-dimensional normal probability distribution. To be more specific, the population Π_1 can be completely specified by the set of $[\frac{r(r+3)}{2}]$ parameters,

$$\{\alpha_{11}, \alpha_{12}, \ldots, \alpha_{1r}; \lambda_{111}, \lambda_{122}, \ldots, \lambda_{1rr}; \lambda_{112}, \lambda_{113}, \ldots, \lambda_{1(r-1)r}\},$$

of which the first r are the location parameters and the remaining indicate the orientation of Π_1. Similarly, Π_2 can also be specified by another set of $[\frac{r(r+3)}{2}]$ parameters,

$$\{\alpha_{21}, \alpha_{22}, \ldots, \alpha_{2r}; \lambda_{211}, \lambda_{222}, \ldots, \lambda_{2rr}; \lambda_{212}, \lambda_{213}, \ldots, \lambda_{2(r-1)r}\},$$

where $(\alpha_{21}, \alpha_{22}, \ldots, \alpha_{2r})$ specify the location and the remaining describe the orientation of Π_2. This shows that the dissimilarity between Π_1 and Π_2 and, for that matter, between any two normal populations can be judged through the disagreement between the two types of parameters; namely, the location and the orientation of these populations. Now, an examination of the expressions (2.5.10) and (2.5.21) for $L(\Pi_1, \Pi_2)$ shows that the first component

measures the dissimilarity of the two populations with respect to their locations, while the second component distinguishes these in terms of their orientations. Thus the function K seems to do the job.

2.5.4 Distance for Assessment of the Appropriateness of a Model

Univariate Case :

Suppose that, to start with, we have n observations, $\underset{\sim}{y} = (y_1, y_2, \ldots, y_n)$, at our disposal. We presume this set of values to be a sample coming from the population $\prod^{(o)}$. Correspondingly, another set $\underset{\sim}{y}^{(u)} = (y_1^{(u)}, y_2^{(u)}, \ldots, y_n^{(u)})$ may be simulated through model $M^{(u)}$ and identified as a sample from the population $\prod^{(u)}$. Now the distributions of $\prod^{(o)}$ and $\prod^{(u)}$ being $N(\alpha^{(o)}, \lambda^{(o)})$ and $N(\alpha^{(u)}, \lambda^{(u)})$, respectively, the dissimilarity between the parent populations of the two samples is given by [Refer equations (2.5.1) and (2.5.2) of Section 2.5]

$$K(\prod^{(o)}, \prod^{(u)}) = [1 - \{\frac{4\,\lambda^{(o)}\lambda^{(u)}}{(\lambda^{(o)} + \lambda^{(u)})^2}\}^{1/4} \exp\{-\frac{(\alpha^{(o)} - \alpha^{(u)})^2}{4(\lambda^{(o)} + \lambda^{(u)})}\}]. \qquad (2.5.24)$$

A sample estimate of $K(\prod^{(o)}, \prod^{(u)})$ and hence of the discrepency of model u in explaining the mechanism of the process may thus be proposed as

$$K_n^{(u)} = [1 - \{\frac{4\, s_n^2\, s_{u,n}^2}{(s_n^2 + s_{u,n}^2)^2}\}^{1/4} \exp\{-\frac{1}{4}\frac{(\bar{y}_n - \bar{y}_{u,n})^2}{(s_n^2 + s_{u,n}^2)}\}]^{1/2}, \qquad (2.5.25)$$

where

$$s_n^2 = \underset{\sim}{y}'(I - \frac{1}{n} J_{nn})\underset{\sim}{y} \ , \ \bar{y}_n = \frac{1}{n} \underset{\sim}{y}' J_{n1}$$

and

$$s_{u,n}^2 = \underset{\sim}{y}^{(u)'} (I - \frac{1}{n} J_{nn})\underset{\sim}{y}^{(u)}, \ \bar{y}_{u,n} = \frac{1}{n} \underset{\sim}{y}^{(u)'} J_{n1}$$

with J_{ab} as an a x b matrix of unit elements. In fact such an estimate is obtained by plugging in the estimates of the unknown quantities, $\alpha^{(o)}$, $\alpha^{(u)}$, $\lambda^{(o)}$, and $\lambda^{(u)}$, involved in (2.5.24) and will be termed as plug-in-estimate.

Multivariate Case : When the system being investigated is multiresponse, the n sets of observations, $\underset{\approx}{y} = (\underset{\sim}{y}_1, \underset{\sim}{y}_2, \ldots, \underset{\sim}{y}_n)'$, arisen from the system may be considered to form a sample from the population $\prod^{(o)}$, while the n sets of simulated observations, $\underset{\approx}{y}^{(u)} = (\underset{\sim}{y}_1^{(u)}, \underset{\sim}{y}_2^{(u)}, \ldots, \underset{\sim}{y}_n^{(u)})'$ may be treated as a sample from $\prod^{(u)}$, u = 1,2,...,m. In this case, since the parent populations have been assumed to have normal distributions, $N_r(\underset{\sim}{\alpha}^{(o)}, \Lambda^{(o)})$ and $N_r(\underset{\sim}{\alpha}^{(u)}, \Lambda^{(u)})$, respectively, the required dissimilarity is given by [Refer equations (2.5.13) and (2.5.14) of Section 2.5]

$$K(\prod^{(o)}, \prod^{(u)}) =$$

$$\left[1 - \left\{\frac{4|\Lambda^{(o)}\Lambda^{(u)}|}{|\Lambda^{(o)}+\Lambda^{(u)}|^2}\right\}^{1/4} \times \right.$$

$$\left. \exp\left\{-\frac{1}{4} (\underset{\sim}{\alpha}^{(o)} - \underset{\sim}{\alpha}^{(u)})'(\Lambda^{(o)} + \Lambda^{(u)})^{-1} (\underset{\sim}{\alpha}^{(o)} - \underset{\sim}{\alpha}^{(u)})\right\}\right]^{1/2}.$$

(2.5.26)

Therefore, a plug-in-estimate for assessing the appropriateness of model u for the process is given by

$$K_n^{(u)} =$$

$$\left[1 - \left\{\frac{4|S_n^{(o)}\ S_n^{(u)}|}{|S_n^{(o)} + S_n^{(u)}|^2}\right\}^{1/4}\right.$$

$$\left. \exp\left\{-\frac{1}{4}\,(\underset{\sim}{\bar{y}}_n^{(o)} - \underset{\sim}{\bar{y}}_n^{(u)})'\ (S_n^{(o)} + S_n^{(u)})^{-1}\ (\underset{\sim}{\bar{y}}_n^{(o)} - \underset{\sim}{\bar{y}}_n^{(u)})\right\}\right]^{1/2}, \qquad (2.5.27)$$

where

$$S_n^{(o)} = \underset{\approx}{y}'\ (I - \tfrac{1}{n}\,J_{nn})\ \underset{\approx}{y}\ ,\quad \underset{\sim}{\bar{y}}_n^{(o)} = \tfrac{1}{n}\,\underset{\approx}{y}'\ J_{n1}$$

and

$$S_n^{(u)} = \underset{\approx}{y}^{(u)'}\ (I - \tfrac{1}{n}\,J_{nn})\underset{\approx}{y}^{(u)}\ ,\quad \underset{\sim}{\bar{y}}_n^{(u)} = \tfrac{1}{n}\,\underset{\approx}{y}^{(u)'}\ J_{n1}$$

Having calculated the estimates of the distances $K(\prod^{(o)}, \prod^{(u)})$ from (2.5.25) or (2.5.27) according as the system is uniresponse or multiresponse, a corresponding change in the discrimination index, defined in (2.4.1), would of course be the next logical step so as to have an over all picture of the discrimination achieved through the available observations at the nth stage.

2.6 TERMINATION CRITERION

In sequential design of experiments for discrimination, at each stage, it is important to assess the level of discrimination achieved upto that stage. It has been proposed

to do so through the discrimination index defined in (2.4.1). Equally important is the decision on termination of the sequential process at the point at which one's goal has been met. Such a decision has considerable significance especially, when the designed experiments are to be conducted on a real life system, where each run of the experiment may be extremely costly. The necessity, therefore, is of a stopping rule through which one could stop at the right stage and avoid unnecessary experimentation.

According to the approach used here, of the two models we prefer the one which has smaller value of the discrimination index, indicating that this model is closer to the true model; the assessment being based on the distance between the corrosponding populations. Thus when m models are being discriminated, a stage would always be awaitted when the models close to the true model corrospond to populations which are sufficiently apart from one another.

In the present approach, the decision on the termination of the sequential procedure can normally be taken on the basis of subjective judgement, i.e., through a comparison of the values attained by the discrimination index at a particular stage. But, there may be situations in which the best model does not differ much from its closest rival, in terms of the distance. In order to tackle these types of situtions we propose to stop at the stage n* (say), if

$$\text{ABS}\left[\frac{D^{(u^+)}_{n^*}}{D^{(u^+)}_{n^*} + D^{(u^*)}_{n^*}} - \frac{D^{(u^+)}_{(n^*-1)}}{D^{(u^+)}_{(n^*-1)} + D^{(u^*)}_{(n^*-1)}}\right] \leq 0.001 , \quad (2.6.1)$$

where $D^{(u^*)}_{(n^*-1)}$, $D^{(u^*)}_{n^*}$ are the values of the discrimination index for model $M^{(u^*)}$ at two consecutive stages and u^*, u^+ stand for the best and the second best models, respectively.

2.6 TESTING THE SIGNIFICANCE OF THE DISTANCE BETWEEN TWO MODELS

In the sense pertinent to this study, the distance between two models, M_1 and M_2 (say), can be measured through the distance between their associated populations, Π_1 and Π_2 (say). Therefore, testing the significance of the distance between two models would amount to confirming whether the two populations, Π_1 and Π_2, are close enough so that the corresponding models, M_1 and M_2, may be considered as substitutes of each other. This can be done by testing the hypothesis,

$$H_0 : K(\Pi_1, \Pi_2) = 0, \qquad (2.7.1)$$

or, equivalently, by testing the hypothesis

$$H_{01} : L(\Pi_1, \Pi_2) = 0, \qquad (2.7.2)$$

where $L(\Pi_1, \Pi_2)$ is given by (2.5.10) or (2.5.21), according as the populations are univarite or multivariate. The equivalence of H_0 and H_{01} follows from the fact that

$$K(\Pi_1, \Pi_2) = 0, \text{ if and only if } [1 - \exp\{-L(\Pi_1, \Pi_2)\}]^{1/2} = 0,$$

i.e., if and only if $[\exp\{-L(\Pi_1, \Pi_2)\}] = 1$,

i.e., if and only if $L(\Pi_1, \Pi_2) = 0$.

ultivariate Case:

We will, therefore, formulate a test for the hypothesis, H_{01}. This will be done for the multivariate case first; the one for the univariate case goes in parallel. A look at equation (2.5.21) further suggests that $L(\Pi_1, \Pi_2) = 0$ if and only if $D_1(\underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2) = 0$ and $D_2(\Lambda_1, \Lambda_2) = 0$. Therefore, the required confirmation about the similarity between Π_1 and Π_2 can be done by testing the ypothesis

$$H_{02} : \underset{\sim}{\alpha}_1 = \underset{\sim}{\alpha}_2 \quad \text{and} \quad \Lambda_1 = \Lambda_2. \tag{2.7.3}$$

To that purpose we consider two hypotheses; namely, $H_1 : \Lambda_1 = \Lambda_2$ and $H_2 : \underset{\sim}{\alpha}_1 = \underset{\sim}{\alpha}_2$, given $\Lambda_1 = \Lambda_2$. Besides, let Ω be the unrestricted parameter space, i.e., $\Omega = \{\underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2, \Lambda_1, \Lambda_2\}$ and $\Omega_1 = \{\underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2, \Lambda_1, \Lambda_2 / \Lambda_1 = \Lambda_2\}$, $\Omega_2 = \{\underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2, \Lambda_1, \Lambda_2\} / \underset{\sim}{\alpha}_1 = \underset{\sim}{\alpha}_2, \Lambda_1 = \Lambda_2\}$ e two restricted parameter spaces. Suppose the r-variate andom vector $\underset{\sim}{Y}$ has p.d.f., $f(\underset{\sim}{y}, \underset{\sim}{\omega})$, where $\underset{\sim}{\omega}$ is a parameter oint in the space Ω. In that case the hypothesis H_1 is that belongs to the space Ω_1; the hypothesis H_2 means that $\underset{\sim}{\omega}$ falls n Ω_2, given that it is in $\Omega_1 \supset \Omega_2$; and finally, H_{02} is the ypothesis that $\underset{\sim}{\omega}$ is a point of Ω_2, given that it is in Ω. uch a formulation enables us to use the lemma due to Anderson 1957). We thus propose to use the statistic, W_1, given by

$$W_1 = \frac{(\nu)^{r\nu/2} \, |\hat{\Lambda}_1|^{\nu_1/2} \, |\hat{\Lambda}_2|^{\nu_2/2}}{|\Lambda^*|^{\nu/2} \, (\nu_1)^{r\nu_1/2} \, (\nu_2)^{r\nu_2/2}}, \tag{2.7.4}$$

or testing the hypothesis under consideration, where ν_u is he number of degrees of freedom associated with the estimate of Λ_u and Λ^* is given by

$$\Lambda^* = \sum_{u=1}^{2} [\hat{\Lambda}_u + n(\bar{\underset{\sim}{y}}_u - \bar{\underset{\sim}{y}})(\bar{\underset{\sim}{y}}_u - \bar{\underset{\sim}{y}})'] \qquad (2.7.5)$$

with $\bar{\underset{\sim}{y}}_u = (\underset{\approx}{y}_u' J_{nn})/n$, $\bar{\underset{\sim}{y}} = (\bar{\underset{\sim}{y}}_1 + \bar{\underset{\sim}{y}}_2)/2$, $u = 1,2$

J_{nn} is the $n \times n$ matrix with all unit elements, and ν in (2.7.4) is the sum, $\nu_1 + \nu_2$.

Using Box (1949) the distribution of W. under H_{02} is given by

$$P(-2\rho \log_e W_1 \leq z) = P(\chi^2_d \leq z) + \gamma\{P(\chi^2_{d+4} \leq z) - P(\chi^2_d \leq z)\} - O(n^{-3}), \qquad (2.7.6)$$

where

$$\rho = 1 - \left(\frac{1}{\nu_1} + \frac{1}{\nu_2} - \frac{1}{\nu}\right)\left[\frac{2r^2+3r-1}{6(r+3)}\right] + \frac{r}{\nu(r+3)} \qquad (2.7.7)$$

$$d = r(r+1)/2 \qquad (2.7.8)$$

$$\gamma = \frac{r}{288\rho^2}\left[6(r+2)(r+1)(r-1)\left(\frac{1}{\nu_1^2} + \frac{1}{\nu_2^2} - \frac{1}{\nu^2}\right)\right.$$
$$- \frac{(2r^2+3r-1)^2}{r+3}\left(\frac{1}{\nu_1} + \frac{1}{\nu_2} - \frac{1}{\nu}\right)^2$$
$$- 12\frac{(2r^2+3r-1)r}{\nu(r+3)}\left(\frac{1}{\nu_1} + \frac{1}{\nu_2} - \frac{1}{\nu}\right)$$
$$\left. - 36\frac{r^2}{(r+3)\nu^2} + 24\frac{r^2-1}{\nu^2}\right], \qquad (2.7.9)$$

and n is the number of observations on the basis of which the decision about the akinness of Π_2 to Π_1 is to be taken.

Under the hypothesis H_{02}, the statistic V, defined as

$$V = -2\rho \log_e W_1 , \qquad (2.7.10)$$

is distributed approximately according to χ^2 distribution with d degrees of freedom. This approximation can be safely used if γ is small, say of the order of 0.001.

Univariate test : A study of the equation (2.5.16), in this case, too, suggests that an equivalent hypothesis would be

$$H_{02} : \alpha_1 = \alpha_2 \text{ and } \lambda_1 = \lambda_2 , \qquad (2.7.11)$$

so that, having defined similar type of parametric spaces and formulating the same set of hypothesis we may propose to utilize the statistic [Anderson (1957)]

$$W_2 = \frac{(\nu_1\hat{\lambda}_1)^{\nu_1/2} (\nu_2\hat{\lambda}_2)^{\nu_2/2}}{[\nu_1\hat{\lambda}_1 + \nu_2\hat{\lambda}_2 + \frac{n_1 n_2}{n_1+n_2} (\hat{\alpha}_1-\hat{\alpha}_2)^2]^{(\nu_1+\nu_2)/2}} \qquad (2.7.12)$$

for testing the given hypothesis, where $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\lambda}_1$, and $\hat{\lambda}_2$ are the appropriate estimates of $\alpha_1, \alpha_2, \lambda_1$, and λ_2, respectively and ν_1, ν_2 are the degrees of freedom associated with $\hat{\lambda}_1$ and $\hat{\lambda}_2$, respectively.

So far as the distribution W_2 is concerned, using Box (1949), the statistic, F, given by

$$F = \frac{d_2 W_2}{(d - W_2)} \qquad (2.7.13)$$

is distributed as Senedecor's F with d_1 and d_2 degrees of freedom, where

$$d_1 = 1 , \; d_2 = 3/\beta^2, \; b = d_2(1 - \beta + 2/d_2)^{-1}, \text{ and}$$

$$\beta = \frac{1}{3} \left[\frac{1}{\nu_1} + \frac{1}{\nu_2} - \frac{1}{\nu_1+\nu_2}\right] .$$

When $\nu_1 = \nu_2$, β assumes the form $(\frac{1}{2\omega})$, so that F of (2.7.13) becomes

$$F = \frac{12\omega^2(6\omega^2 - 3\omega + 1)\, W_2}{[72\omega^4 - (6\omega^2 - 3\omega + 1)W_2]} \qquad (2.7.14)$$

and has an F-distribution with 1 and $(12\omega^2)$ degrees of freedom where $\omega = (\nu_1 + \nu_2)/2$.

2.8 A MODEL ADEQUACY CRITERION

Suppose that in a given situation a single model, u (say), has been chosen from a host of m models through one of the two approaches; namely, sequential and nonsequential. Before this model is finally accepted for actual use it should be examined for its adequacy for the system. According to the concept utilized in this work, the problem of testing the adequacy of a mechanistic model and for that reason of $M^{(u)}$ may be viewed from a different angle. In fact, once we have visualized $\prod^{(o)}$ as the population generated by the system and $\prod^{(u)}$ as the population simulated by the model $M^{(u)}$, this model may be declared adequate for the system if the two populations are sufficiently close to each other, in the prevailing sense of the term. This can be checked by testing the significance of the distance $K(\prod^{(o)}, \prod^{(u)})$, i.e., by testing the hypothesis

$$H_0 : K_n^{(u)} = 0 ,$$

where n denotes the number of observations available upto the stage where such a confirmation is required. As proposed in

Section 2.7 such an hypothesis can be tested through the statistic V, given in (2.7.10) if the model is multivariate. In case the model meant for a single response system is to be tested for adequacy, the univariate analogue of V or the statistic F of equation (2.7.14) may be used. In any case the acceptance of H_0 will lead to the conclusion that the model in question is capable of describing the mechanism of the given system.

CHAPTER 3

DESIGN OF EXPERIMENTS FOR MODEL DISCRIMINATION IN UNIRESPONSE SYSTEMS

3.1 BASIC ASSUMPTIONS

In Chapter 2 we proposed that the criterion function, $\phi(\underset{\sim}{\xi}_{n+1})$, of (2.4.5) could be employed for searching an optimum (from discrimination point of view) set of value(s) of the input variable(s). In the present chapter we will derive particular forms of this function under different sets of assumptions, one may be permitted to make in a given situation while designing experiments for a single response system. Now, there are certain basic assumptions required for the development of the these forms of $\phi(\underset{\sim}{\xi}_{n+1})$. To list these, we consider the model, $M^{(u)}$ (say), i.e., the equation

$$Y_k = \eta^{(u)}(\underset{\sim}{\xi}_k, \underset{\sim}{\theta}^{(u)}) + \varepsilon_k^{(u)} \qquad (3.1.1)$$

and assume that

B(i) $E^{(u)}(\varepsilon_k^{(u)}) = 0$, $k = 1,2,\ldots,n$,

B(ii) $E^{(u)}(\varepsilon_\ell^{(u)}\varepsilon_k^{(u)}) = 0$, $k = 1,2,\ldots,n$; $\ell \neq k$,

B(iii) $E^{(u)}(\varepsilon_k^{(u)2}) = \sigma_u^2$, $k = 1,2,\ldots,n$,

B(iv) the error $\varepsilon_k^{(u)}$ is distributed as $N_1(0, \sigma_u^2)$,

and that

B(v) the model u, if nonlinear, can be linearized in the parameter space;

the assumptions being applicable to all the m rival models.

3.2 CASE 1 : KNOWN, HOMOGENEOUS VARIANCES OF ERRORS

We first consider the case where, in addition to the basic assumptions B(i) through B(v) of Section 3.1, it may be reasonable to assume that

C1(i) $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_m^2 \quad (= \sigma^2)$,

C1(ii) σ^2 is known.

As proposed earlier, the designing of an additional experiment, $\underset{\sim}{\xi}_{n+1}$ (say), requires the use of the criterion function

$$\phi(\underset{\sim}{\xi}_{n+1}) = \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} w_{u,v;n} \, K_{u,v;n+1}.$$

The weights, $w_{u,v;n}$, in this function are supposed to have been obtained through the values of the discrimination index at the nth stage. As regards the specification of the form of $K_{u,v;n+1}$ pertinent to the present set of assumptions, we consider the random variable, Y_{n+1}, and seek its alternative probability distributions under models u and v.

In the present set-up, given $\eta^{(u)}(\underset{\sim}{\xi}_{n+1}, \underset{\sim}{\theta}^{(u)})$ $\quad [= \mu_{n+1}^{(u)} \text{(say)}]$ and σ^2, one can straight write the p.d.f. of Y_{n+1}, under model u, as

$$f^{(u)}(y_{n+1}/\mu_{n+1}^{(u)},\sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\{-\frac{1}{2}(\frac{y_{n+1}-\mu_{n+1}^{(u)}}{\sigma})^2\}. \tag{3.2.1}$$

But, while at the nth stage, $\mu_{n+1}^{(u)}$ is unknown. Therefore, what we actually need is the p.d.f. of Y_{n+1}, given σ^2 only. We may, however, make use of the information available through the previous n observations. This justifies the use of the Bayesian approach.

Now, making use of the assumption B(i) in model (3.1.1), we have

$$E^{(u)}(Y_k) = \eta^{(u)}(\underset{\sim}{\xi}_k, \underset{\sim}{\theta}^{(u)}) . \tag{3.2.2}$$

We will be dealing with a general case if the response function is considered to be nonlinear in parameters. Accordingly, we take $\eta^{(u)}$ as a nonlinear function in terms of $\underset{\sim}{\theta}^{(u)}$. However, the assumption B(v) in that case permits replacement of the response function $\eta^{(u)}(\underset{\sim}{\xi}_k, \underset{\sim}{\theta}^{(u)})$ in (3.2.2) by its linear approximation around some value of $\underset{\sim}{\theta}^{(u)}$ in the parameter space. We choose this value as the m.l.e. $\underset{\sim}{\hat{\theta}}^{(u)}$ of $\underset{\sim}{\theta}^{(u)}$ and assume as well that the estimates, $\hat{\theta}_t^{(u)}$, $t = 1,2,\ldots,p_u$, are fairly close to the model parameters, so that the above approximation may be justified. This enables us to write

$$E^{(u)}(Y_k) = \eta^{(u)}(\underset{\sim}{\xi}_k, \underset{\sim}{\hat{\theta}}^{(u)}) + \sum_{t=1}^{p_u} (\theta_t^{(u)} - \hat{\theta}_t^{(u)}) x_{kt}^{(u)} ,$$

$$k = 1,2,\ldots,n,n+1, \tag{3.2.3}$$

where

$$x_{kt}^{(u)} = \left[\frac{\partial \eta^{(u)}(\underset{\sim}{\xi}_k, \underset{\sim}{\theta}^{(u)})}{\partial \theta_t^{(u)}}\right]_{\underset{\sim}{\theta}^{(u)} = \hat{\underset{\sim}{\theta}}^{(u)}}. \tag{3.2.4}$$

(It may be noted that equation (3.2.3) remains valid even when the response function is linear in parameters.) Alternatively, we can write (3.2.3) as

$$E^{(u)}(Y_k) - \hat{Y}_k^{(u)} = \underset{\sim}{X}_k^{(u)'}(\underset{\sim}{\theta}^{(u)} - \hat{\underset{\sim}{\theta}}^{(u)}), \quad k = 1,2,\ldots,n,n+1,\ldots, \tag{3.2.5}$$

where

$$\hat{y}_k^{(u)} = \eta^{(u)}(\underset{\sim}{\xi}_k, \hat{\underset{\sim}{\theta}}^{(u)}) \tag{3.2.6}$$

and

$$\underset{\sim}{X}_k^{(u)} = (x_{k1}^{(u)}, x_{k2}^{(u)}, \ldots, x_{kp_u}^{(u)})' \tag{3.2.7}$$

with $x_{kt}^{(u)}$ given by (3.2.4).

In particular, we have the identity

$$\mu_{n+1}^{(u)} - \hat{y}_{n+1}^{(u)} = \underset{\sim}{X}_{n+1}^{(u)'}(\underset{\sim}{\theta}^{(u)} - \hat{\underset{\sim}{\theta}}^{(u)}). \tag{3.2.8}$$

Now, the p.d.f. of Y_{n+1}, under model u, given σ^2 and n observations, $\underset{\sim}{y} = (y_1, y_2, \ldots, y_n)$, can be obtained from the formula

$$f_{n+1}^{(u)}(y_{n+1}/\sigma^2) = \int_{R^1} f^{(u)}(y_{n+1}/\mu_{n+1}^{(u)}, \sigma^2) g^{(u)}(\mu_{n+1}^{(u)}/\sigma^2) d\mu_{n+1}^{(u)}. \tag{3.2.9}$$

Of the two densities involved in the integrand, the first is specified in (3.2.1), while the second is not known as yet. However, the relation (3.2.8) suggests that the distribution of

$\mu_{n+1}^{(u)}$ will be the same as that of $\underset{\sim}{X}_{n+1}^{(u)'}\underset{\sim}{\theta}^{(u)}$. The distribution (posterior) of $\underset{\sim}{\theta}^{(u)}$ must, therefore, be obtained, as a first step. To that purpose we consider the likelihood function of the parameters $\underset{\sim}{\theta}^{(u)}$ based on n independent observations, $\underset{\sim}{y}$, i.e., the function

$$L(\underset{\sim}{\theta}^{(u)}/\underset{\sim}{y},\sigma^2) = (2\pi\ \sigma^2)^{-n/2} \exp\ [-\frac{1}{2\sigma^2}\ \underset{\sim}{\varepsilon}^{(u)'}\underset{\sim}{\varepsilon}^{(u)}],$$

where $\underset{\sim}{\varepsilon}^{(u)} = (\varepsilon_1^{(u)},\varepsilon_2^{(u)},\ldots,\varepsilon_n^{(u)})'$ with $\varepsilon_k^{(u)}$ as in (3.1.1). Using the linear form (3.2.5), we have

$$y_k - E^{(u)}(Y_k) = (y_k - \hat{y}_k^{(u)}) - \underset{\sim k}{X}^{(u)'}(\underset{\sim}{\theta}^{(u)} - \underset{\sim}{\hat{\theta}}^{(u)}),$$

i.e.,
$$\varepsilon_k^{(u)} = e_k^{(u)} - \underset{\sim k}{X}^{(u)'}(\underset{\sim}{\theta}^{(u)} - \underset{\sim}{\hat{\theta}}^{(u)}),$$

where
$$e_k^{(u)} = y_k - \hat{y}_k^{(u)}.$$

The likelihood function, therefore, becomes

$$L(\underset{\sim}{\theta}^{(u)}/\underset{\sim}{y},\sigma^2) = (2\pi\ \sigma^2)^{-n/2} \times$$

$$\exp[-\frac{1}{2\sigma^2}\{\underset{\sim}{e}^{(u)} - X^{(u)}(\underset{\sim}{\theta}^{(u)} - \underset{\sim}{\hat{\theta}}^{(u)})\}'$$
$$\{\underset{\sim}{e}^{(u)} - X^{(u)}(\underset{\sim}{\theta}^{(u)} - \underset{\sim}{\hat{\theta}}^{(u)})\}],$$

where $\underset{\sim}{e}^{(u)}$ is an $n \times 1$ vector of the discrepencies, $e_k^{(u)}$, and $X^{(u)}$ is an $n \times p_u$ matrix of the partial derivatives, $x_{kt}^{(u)}$, $k = 1,2,\ldots,n$; $t = 1,2,\ldots,p_u$, defined in (3.2.4). Now the use of m.l.e. $\underset{\sim}{\hat{\theta}}^{(u)}$ renders $\underset{\sim}{e}^{(u)'}X^{(u)} = 0$ which in turn reduces the likelihood function to the form

$$L(\underset{\sim}{\theta}^{(u)}/\underset{\sim}{y},\sigma^2) =$$

$$(2\pi\,\sigma^2)^{-n/2} \times$$

$$\exp\left[-\frac{1}{2\sigma^2}\{\underset{\sim}{e}^{(u)'}\underset{\sim}{e}^{(u)}+(\underset{\sim}{\theta}^{(u)}-\hat{\underset{\sim}{\theta}}^{(u)})'X^{(u)'}X^{(u)}(\underset{\sim}{\theta}^{(u)}-\hat{\underset{\sim}{\theta}}^{(u)})\}\right]. \tag{3.2.10}$$

The ignorance about the model parameters can be taken into account by assuming uniform prior distribution of $\underset{\sim}{\theta}^{(u)}$. The posterior density of $\underset{\sim}{\theta}^{(u)}$ is, then, given by

$$g(\underset{\sim}{\theta}^{(u)}/\underset{\sim}{y},\sigma^2) = \frac{L(\underset{\sim}{\theta}^{(u)}/\underset{\sim}{y},\sigma^2)}{\int_{R_\theta} L(\underset{\sim}{\theta}^{(u)})/\underset{\sim}{y},\sigma^2)d\underset{\sim}{\theta}^{(u)}}\,, \tag{3.2.11}$$

where R_θ denotes the parameter space for model u. The use of L from (3.2.10) in the formula (3.2.11), immediately gives

$$g(\underset{\sim}{\theta}^{(u)}/\underset{\sim}{y},\sigma^2) = [(2\pi\,\sigma^2)^{p_u}\,|X^{(u)'}X^{(u)}|]^{-1/2} \times$$

$$\exp\left[-\frac{1}{2\sigma^2}\{(\underset{\sim}{\theta}^{(u)}-\hat{\underset{\sim}{\theta}}^{(u)})'X^{(u)'}X^{(u)}(\underset{\sim}{\theta}^{(u)}-\hat{\underset{\sim}{\theta}}^{(u)})\}\right]. \tag{3.2.12}$$

This shows that $\underset{\sim}{\theta}^{(u)}$ is distributed as $N_{p_u}(\hat{\underset{\sim}{\theta}}^{(u)}, (X^{(u)'}X^{(u)})^{-1}\sigma^2)$. So from the relation (3.2.8) we conclude that $\mu_{n+1}^{(u)}$ is distributed normally with mean $\hat{y}_{n+1}^{(u)}$ and variance $[\underset{\sim}{X}_{n+1}^{(u)'}(X^{(u)'}X^{(u)})^{-1}\,\underset{\sim}{X}_{n+1}^{(u)}\,\sigma^2]$. The p.d.f. of $\mu_{n+1}^{(u)}$, given σ^2, can accordingly be written as

$$g^{(u)}(\mu_{n+1}^{(u)}/\sigma^2) = (2\pi\,\sigma^2 z_u)^{-1/2}\exp\left[-\frac{1}{2\sigma^2 z_u}\{(\mu_{n+1}^{(u)} - \hat{y}_{n+1}^{(u)})^2\}\right], \tag{3.2.13}$$

where

$$z_u = \underset{\sim}{x}^{(u)}_{n+1}{}' (X^{(u)}{}' X^{(u)})^{-1} \underset{\sim}{x}^{(u)}_{n+1}. \qquad (3.2.14)$$

Substituting for $f^{(u)}(y_{n+1}/\mu^{(u)}_{n+1}, \sigma^2)$ from (3.2.1) and for $g^{(u)}(\mu^{(u)}_{n+1}/\sigma^2)$ from (3.2.13) in the formula (3.2.9), we get

$$f^{(u)}_{n+1}(y_{n+1}/\sigma^2) =$$

$$[(2\pi\ \sigma^2)^2\ z_u]^{-1/2}$$

$$\int_{R^1} \exp\ [-\tfrac{1}{2}\ \{\frac{(\mu^{(u)}_{n+1} - y_{n+1})^2}{\sigma^2} + \frac{(\mu^{(u)}_{n+1} - \hat{y}^{(u)}_{n+1})^2}{\sigma^2\ z_u}\}]\ d\mu^{(u)}_{n+1}\ . \qquad (3.2.15)$$

Using (A.1.11) and (A.1.12) of Corollary A.1 (Appendix A), we can write

$$[\frac{(\mu^{(u)}_{n+1} - y_{n+1})^2}{\sigma^2} + \frac{(\mu^{(u)}_{n+1} - \hat{y}^{(u)}_{n+1})^2}{\sigma^2 z_u}] =$$

$$\frac{\omega_u^2}{\sigma^4 z_u}\ (\mu^{(u)}_{n+1} - \tilde{\mu})^2 + \frac{(y_{n+1} - \hat{y}^{(u)}_{n+1})^2}{\omega_u^2}\ ,$$

where $\tilde{\mu} = [\sigma^2(z_u y_{n+1} + \hat{y}^{(u)}_{n+1})/\omega_u^2]$

and

$$\omega_u = \sigma(1 + z_u)^{1/2}\ . \qquad (3.2.16)$$

As a result, from (3.2.15), we get

$$f^{(u)}_{n+1}(y_{n+1}/\sigma^2) = (2\pi\ \omega_u^2)^{-1/2} \exp\ [-\tfrac{1}{2}\ (\frac{y_{n+1} - \hat{y}^{(u)}_{n+1}}{\omega_u})^2]\ . \qquad (3.2.17)$$

This leads to the conclusion that, under model u, Y_{n+1}, in the present set-up, is distributed as $N_1(\hat{y}^{(u)}_{n+1}, \omega_u^2)$. Similarly, under model v($v \neq u$), Y_{n+1} will be distributed as $N_1(\hat{y}^{(u)}_{n+1}, \omega_v^2)$ so that

the p.d.f. of Y_{n+1}, under model v, given σ^2 and n observation, can be written as

$$f_{n+1}^{(v)}(y_{n+1}/\sigma^2) = (2\pi\,\omega_v^2)^{-1/2} \exp\left[-\frac{1}{2}\left(\frac{y_{n+1} - \hat{y}_{n+1}^{(v)}}{\omega_v}\right)^2\right], \qquad (3.2.18)$$

where $\omega_v = \sigma(1 + z_v)^{1/2}$.

Using (3.2.17) and (3.2.18) in the relation

$$h_1(f_{n+1}^{(u)}, f_{n+1}^{(v)}) = \int_{R^1} [f_{n+1}^{(u)}(y_{n+1}/\sigma^2) f_{n+1}^{(v)}(y_{n+1}/\sigma^2)\,]^{1/2} dy_{n+1},$$

we get

$$h_1(f_{n+1}^{(u)}, f_{n+1}^{(v)}) = [(2\pi)^2\,\omega_u^2\omega_v^2]^{-1/4} \int_{R^1} \exp\left[-\frac{1}{4}\left\{\frac{(y_{n+1} - \hat{y}_{n+1}^{(u)})^2}{\omega_u^2} + \frac{(y_{n+1} - \hat{y}_{n+1}^{(u)})^2}{\omega_v^2}\right\}\right] dy_{n+1}\,. \qquad (3.2.19)$$

Now by Corollary A.1 of Appendix A we can write

$$\frac{(y_{n+1} - \hat{y}_{n+1}^{(u)})^2}{\omega_u^2} + \frac{(y_{n+1} - \hat{y}_{n+1}^{(v)})^2}{\omega_v^2} = \frac{(y_{n+1} - \tilde{y}_{n+1})^2}{\tilde{\omega}^2} + \frac{(\hat{y}_{n+1}^{(u)} - \hat{y}_{n+1}^{(v)})^2}{\omega_u^2 + \omega_v^2}, \qquad (3.2.20)$$

where $\tilde{y}_{n+1} = (\omega_v^2\,\hat{y}_{n+1}^{(u)} + \omega_u^2\,\hat{y}_{n+1}^{(v)})/(\omega_u^2 + \omega_v^2)$ and

$$\tilde{\omega} = \{\omega_u^2\,\omega_v^2/(\omega_u^2 + \omega_v^2)\}^{1/2}\,. \qquad (3.2.21)$$

Using (3.2.20)in (3.2.19), we get

$$h_1(f_{n+1}^{(u)}, f_{n+1}^{(v)}) = [(2\pi)^2\, \omega_u^2\omega_v^2]^{-1/4} \exp\left[-\frac{1}{4}\left\{\frac{(\hat{y}_{n+1}^{(u)} - \hat{y}_{n+1}^{(v)})^2}{\omega_u^2 + \omega_v^2}\right\}\right]$$

$$\int_{R^1} \exp\left[-\frac{1}{4}\left\{\frac{(y_{n+1} - \tilde{y}_{n+1})^2}{\tilde{\omega}^2}\right\}\right] dy_{n+1} ,$$

i.e.,

$$h_1(f_{n+1}^{(u)}, f_{n+1}^{(v)}) = [(2\tilde{\omega})^{-2}\, \omega_u^2\omega_v^2]^{-1/4} \exp\left[-\frac{1}{4}\left\{\frac{(\hat{y}_{n+1}^{(u)} - \hat{y}_{n+1}^{(v)})^2}{\omega_u^2 + \omega_v^2}\right\}\right]$$

Substituting for $\tilde{\omega}$ from (3.2.20) we finally get

$$h_1(f_{n+1}^{(u)}, f_{n+1}^{(v)}) = \left[\frac{4\omega_u^2\omega_v^2}{(\omega_u^2 + \omega_v^2)^2}\right]^{1/4} \exp\left[-\frac{1}{4}\left\{\frac{(\hat{y}_{n+1}^{(u)} - \hat{y}_{n+1}^{(v)})^2}{\omega_u^2 + \omega_v^2}\right\}\right]. \tag{3.2.22}$$

We conclude this section by proposing that the appropriate design criterion, when σ is known and the models are linear or can be linearized, consists of maximizing the function

$$\phi\,(\underset{\sim}{\xi}_{n+1}) =$$

$$\sum_{u=1}^{m-1} \sum_{v=u+1}^{m} w_{u,v;n} \left[1 - \left(\frac{2\omega_u\omega_v}{\omega_u^2 + \omega_v^2}\right)^{1/2} \exp\left\{-\frac{1}{4}\left(\frac{(\hat{y}_{n+1}^{(u)} - \hat{y}_{n+1}^{(v)})^2}{\omega_u^2 + \omega_v^2}\right)\right\}\right]^{1/2} \tag{3.2.23}$$

with respect to $\underset{\sim}{\xi}_{n+1}$, within the operability region.

3.3 CASE 2 : UNKNOWN, HOMOGENEOUS VARIANCES OF ERRORS

In actual practice, the information about the variance of errors is often lacking. However, the normality and additivity

of errors as well as the linearization of nonlinear models may still be possible. In the present case, therefore, we retain all the assumptions B(i) through B(v) of Section 3.1. In addition, we assume that

C2(i) $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_m^2 \quad (= \sigma^2)$,

C2(ii) σ^2 is unknown.

Now in order to be able to use the objective function (2.4.5) for designing new experiments we derive the form of $K_{u,v;n+1}$, relevant to the present case. To that purpose, we first consider a random variable, Y_{n+1}, and seek its p.d.f. (with reference to future), under models u and v, when all the above mentioned assumptions hold good. Of all these, if the assumption C2(ii) is suppressed for the time being, then through the discussion in Section 3.2 we know that the p.d.f. of Y_{n+1}, under model u, given σ^2 $(= \lambda(\text{say}))$ can be written as [Refer equation (3.2.17)]

$$f_{n+1}^{(u)}(y_{n+1}/\lambda) = [2\pi\,\lambda(1+z_u)]^{-1/2} \exp\left[-\frac{1}{2}\left\{\frac{(y_{n+1} - \hat{y}_{n+1}^{(u)})^2}{\lambda(1+z_u)}\right\}\right], \tag{3.3.1}$$

where $\underset{\sim}{y}$ is the set of previous n observations and

$$z_u = \underset{\sim}{X}_{n+1}^{(u)\prime} [X^{(u)\prime} X^{(u)}]^{-1} \underset{\sim}{X}_{n+1}^{(u)} , \tag{3.3.2}$$

while the vector $\underset{\sim}{X}_{n+1}^{(u)}$ and the matrix $X^{(u)}$ are defined earlier in Section 3.2.

Since in the situation under consideration λ is unknown, we use some appropriate prior for λ and consider the marginal p.d.f. of Y_{n+1}, i.e.,

$$f^{(u)}_{n+1}(y_{n+1}/\underset{\sim}{y}) = \int_0^\infty f^{(u)}(y_{n+1}/\lambda)\ g(\lambda/\underset{\sim}{y})\ d\lambda\ , \qquad (3.3.3)$$

where $g(\lambda/\underset{\sim}{y})$ is the posterior density of λ, given a sample of n observations. We use a noninformative prior for λ. According to Jeffrey's (1961) rule, the prior distribution of a parameter, θ^* (say), is approximately noninformative if it is taken proportional to the square root of its Fisher's information measure, $\mathcal{J}(\theta^*)$. Since in the present case such a measure is given by

$$\mathcal{J}(\lambda) \propto 1/\lambda^2,$$

the desired prior can be taken as

$$g(\lambda) \propto 1/\lambda\ . \qquad (3.3.4)$$

Besides, the sampling distribution of the sum of squares, $(\nu\hat{\sigma}^2/\lambda)$, is a Chi-square distribution with ν degrees of freedom. Consequently, we have

$$g(\hat{\sigma}^2/\lambda) = \frac{(\nu/2\lambda)^{\nu/2}}{\Gamma(\frac{\nu}{2})}\ (\hat{\sigma}^2)^{\frac{\nu}{2}-1}\ \exp\ (-\frac{\nu\hat{\sigma}^2}{2\lambda}),\ \lambda > 0, \qquad (3.3.5)$$

where $\hat{\sigma}^2$ is an appropriate estimate of σ^2, ν is the number of degrees of freedom associated with this estimate and Γ denotes the gamma function, defined in Appendix D. Since in the present case $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_m^2\ (= \sigma^2)$, the estimate $\hat{\sigma}^2$ is, in fact, obtained by pooling the estimates, $\hat{\sigma}_u^2$'s of σ_u^2's, each

secured through the corresponding model. Thus

$$\hat{\sigma}^2 = \frac{\sum_{u=1}^{m} \nu_u \hat{\sigma}_u^2}{\nu}, \qquad (3.3.6)$$

where $\nu = \sum_{u=1}^{m} \nu_u$ and ν_u is the number of degrees of freedom due to the estimate $\hat{\sigma}_u^2$.

So far as the posterior, $g(\lambda/\underset{\sim}{y})$ $[= g(\lambda/\hat{\sigma}^2)]$, of λ is concerned, we may use (3.3.4) and (3.3.5) in the relation

$$g(\lambda/\hat{\sigma}^2) \propto g(\hat{\sigma}^2/\lambda)\, g(\lambda).$$

Keeping in view the fact that for a given sample of n observations the quantity $[\{(\nu/2)^{\nu/2}/\Gamma(\nu/2)\}(\hat{\sigma}^2)^{(\nu/2-1)}]$, is a fixed constant, we can write

$$g(\lambda/\underset{\sim}{y}) = C\, \lambda^{-(\frac{\nu}{2}+1)} \exp\{-(\frac{\hat{\nu}\sigma^2}{2})\lambda^{-1}\},$$

where

$$C^{-1} = \int_0^\infty \lambda^{-(\frac{\nu}{2}+1)} \exp\{-(\frac{\nu\hat{\sigma}^2}{2})\lambda^{-1}\}\, d\lambda$$

Using the inverse gamma function (See Appendix D), we get

$$C^{-1} = (\frac{\nu\hat{\sigma}^2}{2})^{-\nu/2}\, \Gamma(\frac{\nu}{2}).$$

Thus the posterior p.d.f. of λ finally reduces to the form

$$g(\lambda/\underset{\sim}{y}) = \frac{(\frac{\nu\hat{\sigma}^2}{2})^{\nu/2}}{\Gamma(\frac{\nu}{2})} \lambda^{-(\frac{\nu}{2}+1)} \exp\{-(\frac{\nu\hat{\sigma}^2}{2})\lambda^{-1}\}, \quad \lambda > 0 \qquad (3.3.7)$$

We now utilize (3.3.1) and (3.3.7) in the formula (3.3.3) and obtain

$$f_{n+1}^{(u)}(y_{n+1}/\underset{\sim}{y}) =$$

$$\left(\frac{\nu\hat{\sigma}^2}{2}\right)^{\nu/2} \left[\{2\pi(1+z_u)\}^{1/2} \ \Gamma(\nu/2)\right]^{-1} \times$$

$$\int_0^\infty \lambda^{-\left(\frac{\nu+1}{2}+1\right)} \exp\left[-\frac{1}{2}\left\{\nu\hat{\sigma}^2 + \frac{(y_{n+1} - \hat{y}_{n+1}^{(u)})^2}{(1+z_u)}\right\}\lambda^{-1}\right] d\lambda \ . \tag{3.3.8}$$

The integral on the right hand side of (3.3.8) can be solved by the inverse gamma function (See Appendix D) so that

$$f_{n+1}^{(u)}(y_{n+1}/\underset{\sim}{y}) =$$

$$\left(\frac{\nu\hat{\sigma}^2}{2}\right)^{\nu/2} \left[\ \Gamma(1/2)\{2(1+z_u)\}^{1/2} \ \Gamma(\nu/2)\right]^{-1} \times$$

$$\left[\frac{1}{2}\left\{\nu\hat{\sigma}^2 + \frac{(y_{n+1} - \hat{y}_{n+1}^{(u)})^2}{1+z_u}\right\}\right]^{-(\nu+1)/2} \Gamma\left(\frac{\nu+1}{2}\right)$$

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \ \Gamma\left(\frac{1}{2}\right)} \left[\nu\hat{\sigma}^2(1+z_u)\right]^{-1/2} \left[1 + \frac{(y_{n+1} - \hat{y}_{n+1}^{(u)})^2}{\nu\hat{\sigma}^2(1+z_u)}\right]^{-(\nu+1)/2} .$$

The p.d.f. of Y_{n+1}, given n observations, $\underset{\sim}{y}$, is thus given by

$$f_{n+1}^{(u)}(y_{n+1}/\underset{\sim}{y}) = \left[B(\nu/2,1/2) \ \nu^{1/2} \ \hat{\omega}_u\right]^{-1} \left[1 + \frac{(y_{n+1} - \hat{y}_{n+1}^{(u)})^2}{\nu\hat{\omega}_u^2}\right]^{-(\nu+1)/2}, \tag{3.3.9}$$

where

$$\hat{\omega}_u = \hat{\sigma}(1 + z_u)^{1/2} \tag{3.3.10}$$

and B(p,q) stands for the Beta function, given in D.2 of Appendix D. Similarly, the p.d.f. of Y_{n+1}, under model v, in the present set of assumptions is given by

$$f^{(v)}_{n+1}(y_{n+1}/\underset{\sim}{y}) = [B(\nu/2,1/2)\ \nu^{1/2}\ \hat{\omega}_v]^{-1}[1 + \frac{(y_{n+1} - \hat{y}^{(v)}_{n+1})^2}{\nu\hat{\omega}^2_v}]^{-(\nu+1)/2}, \tag{3.3.11}$$

where

$$\hat{\omega}_v = \hat{\sigma}(1 + z_v)^{1/2}. \tag{3.3.12}$$

The densities specified in (3.3.9) and (3.3.11), when used in the relation

$$h_2(f^{(u)}_{n+1}, f^{(v)}_{n+1}) = \int_{R^1} [f^{(u)}_{n+1}(y_{n+1}/\underset{\sim}{y})\ f^{(v)}_{n+1}(y_{n+1}/\underset{\sim}{y})]^{1/2}\ dy_{n+1}\ ,$$

give

$$h_2(f^{(u)}_{n+1}, f^{(v)}_{n+1}) = [B(\nu/2,1/2)\ \nu^{1/2}\ (\hat{\omega}^2_u\ \hat{\omega}^2_v)^{1/4}]^{-1} \times$$

$$\int_{R^1} (1 + \frac{t_u}{\nu})^{-(\nu+1)/4} (1 + \frac{t_v}{\nu})^{-(\nu+1)/4}\ dy_{n+1}, \tag{3.3.13}$$

where

$$t_i = [(y_{n+1} - \hat{y}^{(i)}_{n+1})^2/\hat{\omega}^2_i],\ i = u,v\ . \tag{3.3.14}$$

The integrand in (3.3.13), however, needs to be reduced to a tractable form for integration. To that purpose we write

$$(1 + \frac{t_u}{\nu})^{-(\nu+1)/4} = \exp\{-\frac{\nu+1}{4} \log_e (1 + \frac{t_u}{\nu})\} \qquad (3.3.15)$$

If we use the expansion of $\log_e(1 + t_u\nu^{-1})$, in terms of $(t_u\nu^{-1})$, in the component $[\{(\nu+1)/4\} \log_e(1 + t_u\nu^{-1})]$ and arrange the terms in the ascending order of the powers of ν^{-1} we can express this component as

$$(\frac{\nu+1}{4}) \log_e(1 + \frac{t_u}{\nu}) = \frac{1}{4} t_u - \sum_{i=1}^{\infty} \tau_i \nu^{-i}, \qquad (3.3.16)$$

where τ_i is related to t_u through the expression,

$$\tau_i = \frac{(-1)^i}{4i(i+1)} \{(i+1)\, t_u^i - i\, t_u^{i+1}\}. \qquad (3.3.17)$$

(Refer D.6 of Appendix D for details of the algebra used).

Using (3.3.16), the equation (3.3.15) can thus be reformed as

$$\begin{aligned}(1 + \frac{t_u}{\nu})^{-(\nu+1)/4} &= \exp(-\frac{1}{4} t_u) \exp(\sum_{i=1}^{\infty} \tau_i \nu^{-i}) \\ &= \exp(-\frac{1}{4} t_u) [\prod_{i=1}^{\infty} \exp(\tau_i \nu^{-i})] .\end{aligned} \qquad (3.3.18)$$

We can further simplify the right hand side of this equation by using the expansions of the factors, $\exp(\tau_i\nu^{-i})$. In fact, when the coefficients of the like powers of ν^{-1} are collected in the infinite product of these expansions, we get

$$\prod_{i=1}^{\infty} \exp(\tau_i \nu^{-i}) = \sum_{i=0}^{\infty} a_i \nu^{-i}, \qquad (3.3.19)$$

where

$$a_o = 1 ,$$

$$a_1 = \tau_1 ,$$

$$a_2 = \frac{\tau_1^2}{2} + \tau_2 ,$$

$$a_3 = \frac{\tau_1^3}{2} + \tau_1\tau_2 + \tau_3 ,$$

$$a_4 = \frac{\tau_1^4}{24} + \frac{\tau_1^2\tau_2}{2} + \frac{\tau_2^2}{2} + \tau_1\tau_3 + \tau_4 ,$$

$$a_5 = \frac{\tau_1^5}{120} + \frac{\tau_1^3\tau_2}{6} + \frac{\tau_1\tau_2^2}{2} + \frac{\tau_1^2\tau_3}{2} + \tau_1\tau_4 + \tau_2\tau_3 + \tau_5 ,$$

$$a_6 = \frac{\tau_1^6}{720} + \frac{\tau_1^4\tau_2}{24} + \frac{\tau_1^4\tau_4}{2} + \frac{\tau_1^2\tau_2^2}{4} + \frac{\tau_1^3\tau_3}{6} + \frac{\tau_2^3}{6} + \tau_1\tau_2\tau_3 + \frac{\tau_3^2}{2} + \tau_1\tau_5 + \tau_2\tau_4 + \tau_6 ,$$

etc.

Further, by utilizing the relation (3.3.17), between τ_i and t_u, in the above equations we can express each coefficient of the series in (3.3.19) as a polynomial in t_u. Let $A_i(t_u)$ stand for the polynomial corresponding to a_i. Then

$$A_i(t_u) = \sum_{p=1}^{i+1} \alpha_{ip}\, t_u^{2i-p+1} , \qquad (3.3.20)$$

where α_{ip}'s are given in Table 3.3.1.

Table 3.3.1
Values of the coefficients, α_{ip}, β_{jq}.

i,J \ P,Q	1	2	3	4	5	6	7
0	1.000						
1	0.125	-0.250					
2	0.781E-2	-0.115	0.156				
3	0.326E-3	-0.124E-1	0.911E-2	-0.117			
4	0.102E-4	-0.732E-3	0.015E-1	-0.933E-1	0.952		
5	0.300E-6	-0.297E-4	0.114E-2	-0.168E-1	0.856E-1	-0.899E-1	
6	0.530E-8	-0.911E-6	0.558E-4	-0.737E-2	0.179E-1	-0.793E-1	0.708E-1

E-K stands for $\times 10^{-k}$

A corresponding change in (3.3.19) gives

$$\prod_{i=1}^{\infty} \exp\,(\tau_i \nu^{-i}) = \sum_{i=0}^{\infty} A_i(t_u)\, \nu^{-i} \,. \qquad (3.3.21)$$

Accordingly, when we use (3.3.21) in (3.3.18) we get

$$(1 + \frac{t_u}{\nu})^{-(\nu+1)/4} = \exp\,(-\tfrac{1}{4}\, t_u)\, [\sum_{i=0}^{\infty} A_i(t_u)\nu^{-i}] \,. \qquad (3.3.22)$$

Proceeding on the same lines, we can express the second factor of the integrand in (3.3.13) as

$$(1 + \frac{t_v}{\nu})^{-(\nu+1)/4} = \exp(-\tfrac{1}{4}\, t_v)\, [\sum_{i=0}^{\infty} B_j(t_v)\, \nu^{-j}], \qquad (3.3.23)$$

where $B_j(t_v)$, like $A_i(t_u)$, is a polynomial in t_v and is given by

$$B_j(t_v) = \sum_{q=1}^{j+1} \beta_{jq}\, t_v^{2j-q+1} \qquad (3.3.24)$$

with β_{jq}'s as given in Table 3.3.1.

When we make use of (3.3.22) and (3.3.23), the integrand in (3.3.13) becomes

$$(1 + \frac{t_u}{\nu})^{-(\nu+1)/4} (1 + \frac{t_v}{\nu})^{-(\nu+1)/4} =$$

$$[\exp\,\{-\tfrac{1}{4}\,(t_u + t_v)\}] \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} Q_{ij}(t_u, t_v)\, \nu^{-(i+j)}, \qquad (3.3.25)$$

where

$$Q_{ij}(t_u, t_v) = \sum_{p=1}^{i+1} \sum_{q=1}^{j+1} \alpha_{ip}\beta_{jq}\, t_u^{2i-p+1}\, t_v^{2j-q+1} \qquad (3.3.26)$$

is a bivariate polynomial in t_u and t_v, obtained through the

product, $[A_i(t_u)B_j(t_v)]$, of the polynomials in (3.3.20) and (3.3.24). Further, recalling the definition of t_i $(i = u,v)$ from (3.3.14), we have

$$t_u + t_v = \frac{(y_{n+1} - \hat{y}^{(u)}_{n+1})^2}{\hat{\omega}_u^2} + \frac{(y_{n+1} - \hat{y}^{(v)}_{n+1})^2}{\hat{\omega}_v^2} \quad . \tag{3.3.27}$$

However, through the application of the results (A.1.11) and (A.1.12) [Corollary A.1, Appendix A] on the right hand side of (3.3.27) we can express $(t_u + t_v)$, alternatively, as

$$t_u + t_v = \frac{(y_{n+1} - \tilde{y}_{n+1})^2}{\tilde{\omega}^2} + \frac{(y^{(u)}_{n+1} - y^{(v)}_{n+1})^2}{\omega_u^2 + \omega_v^2} \quad , \tag{3.3.28}$$

where $\tilde{y}_{n+1} = (\hat{\omega}_v^2 \hat{y}^{(u)}_{n+1} + \hat{\omega}_u^2 \hat{y}^{(v)}_{n+1})/(\hat{\omega}_u^2 + \hat{\omega}_v^2)$ and

$$\tilde{\omega} = \left(\frac{\hat{\omega}_u^2 \hat{\omega}_v^2}{\hat{\omega}_u^2 + \hat{\omega}_v^2}\right)^{1/2} \quad . \tag{3.3.29}$$

Substituting (3.3.28) in (3.3.25), we get

$$\left(1 + \frac{t_u}{\nu}\right)^{-(\nu+1)/4} \left(1 + \frac{t_v}{\nu}\right)^{-(\nu+1)/4} =$$

$$\exp\left[-\left\{\frac{(\hat{y}^{(u)}_{n+1} - \hat{y}^{(v)}_{n+1})}{\hat{\omega}_u^2 + \hat{\omega}_v^2}\right\}\right] \left[\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} Q_{ij}(t_u, t_v)\right] \nu^{-(i+j)} \times$$

$$\exp\left[-\frac{1}{4}\left\{\frac{(y_{n+1} - \tilde{y}_{n+1})^2}{\tilde{\omega}^2}\right\}\right] \quad . \tag{3.3.30}$$

When the expression, from (3.3.30), for the integrand is introduced in (3.3.13) h_2 reduces to the form

$$h_2(f^{(u)}_{n+1}, f^{(v)}_{n+1}) =$$

$$[\, B(\nu/2,1/2)\ \nu^{1/2}\ (\hat{\omega}_u^2\ \hat{\omega}_v^2)^{1/4}]^{-1}\ \exp\ [-\tfrac{1}{4}\ \{\frac{(\hat{y}^{(u)}_{n+1} - \hat{y}^{(v)}_{n+1})^2}{\hat{\omega}_u^2 + \hat{\omega}_v^2}\}]\ x$$

$$\int_{R^1} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} Q_{ij}(t_u,t_v)\ \nu^{-(i+j)}\ \exp[-\tfrac{1}{2}\ \{\frac{(y_{n+1} - \tilde{y}_{n+1})^2}{2\tilde{\omega}^2}\}]dy_{n+1}\ .$$

(3.3.31)

Now, the integral in (3.3.31) can be evaluated term by term. In this connection, the integrand involved therein suggests that if Y_{n+1} is considered to be distributed as $N_1(\tilde{y}_{n+1}, 2\tilde{\omega}^2)$, then each integral, involving the bivariate polynomial Q_{ij} would yield a polynomial of the mixed moments of t_u and t_v, of the type $\mu_{a,b} = E(t_u^a t_v^b)$. Using these facts in (3.3.31) we get

$$h_2(f^{(u)}_{n+1}, f^{(v)}_{n+1}) = [\, B(\nu/2,1/2)\nu^{1/2}(\hat{\omega}_u^2\hat{\omega}_v^2)^{1/4}]^{-1}\ [2\pi(2\tilde{\omega}^2)]^{1/2}\ x$$

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{p=1}^{i+1} \sum_{q=1}^{j+1} \alpha_{ip}\beta_{jq}\ \mu_{2i-p+1,2j-q+1}\ \nu^{-(i+j)}\ x$$

$$\exp\ [-\tfrac{1}{4}\ \{\frac{(\hat{y}^{(u)}_{n+1} - \hat{y}^{(v)}_{n+1})^2}{\hat{\omega}_u^2 + \hat{\omega}_v^2}\}].$$

Finally, when we use the expression for $\tilde{\omega}$ from (3.3.29) and rearrange the terms we are able to express h_2 as

$$h_2(f^{(u)}_{n+1}, f^{(v)}_{n+1}) = C^*[\frac{4\hat{\omega}_u^2\ \hat{\omega}_v^2}{(\hat{\omega}_u^2 + \hat{\omega}_v^2)^2}]^{1/4}\ \exp[-\tfrac{1}{4}\ \{\frac{(\hat{y}^{(u)}_{n+1} - \hat{y}^{(v)}_{n+1})^2}{\hat{\omega}_u^2 + \hat{\omega}_v^2}\}],$$

(3.3.32)

where

$$C^* = [2\pi/\ B(\nu/2,1/2)]^{1/2} \times$$

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{p=1}^{i+1} \sum_{q=1}^{j+1} \alpha_{ip}\beta_{jq}\, \mu_{2i-p+1,2j-q+1}\, \nu^{-(i+j+\frac{1}{2})} . \qquad (3.3.33)$$

Substitution of h_2 from (3.3.32) for h in the function

$$K(f_{n+1}^{(u)}, f_{n+1}^{(v)}) = [1 - h(f_{n+1}^{(u)}, f_{n+1}^{(v)}]^{1/2}$$

would give the required expression for $K_{u,v}(\underset{\sim}{\xi}_{n+1})$ and hence the criterion function $\phi\ (\underset{\sim}{\xi}_{n+1})$ for the case when the common variance of error, σ^2, is unknown and the information available from the previous n observations is utilized.

The mixed moments, $\mu_{a,b}$, involved in (3.3.33) are difficult to evaluate. However, if we make use of the inversion formulae of Cook (1951), given in D.5 of Appendix D, then we can find at least some of these, conveniently, through the joint cumulants k_{ab}. Fortunately, the number of these formulae happens to be adequate for our purpose as will be argued latter. So far as the evaluation of k_{ab} is concerned, the fact that Y_{n+1} may be considered to be distributed as $N_1(\tilde{y}_{n+1}, \tilde{\omega})$ enables us to use Lemma C.1 of Appendix C and get the following expressions for the required cumulants

$$k_{a0} = 2^{a-1}(a-1)!\ (\rho_u/\rho)^a\ [1 + a\,\rho\, d_u^2] \ ,$$

$$k_{0b} = 2^{b-1}(b-1)!\ (\rho_v/\rho)^b\ [1 + b\,\rho\, d_v^2] \ ,$$

$$k_{ab} = 2^{a+b-1}(a+b-2)!\ (\rho_u/\rho)^a(\rho_v/\rho)^b \times$$

$$[(a+b-1) + a(a-1)\,\rho\, d_u^2 + b(b-1)\rho\ d_v^2 + 2ab\,\rho\, d_u d_v],$$

where $\rho_s = (\hat{\omega}_s)^{-1}$, $d_s = (\tilde{y}_{n+1} - \hat{y}_{n+1}^{(s)})$, $s = u,v$ and $\rho = (\rho_u + \rho_v)/2$.

Remarks :

(i) A close look at the expression in (3.3.32) brings out the important point that the function h_2 and hence $K_{u,v}(\underset{\sim}{\xi}_{n+1})$ takes into account the dissimilarity which a new observation is likely to introduce, in terms of the expectations and orientations of the underlying probability distributions associated with models u and v.

(ii) The equation (3.3.32) also points out that, except for the factor C^*, the expression for h_2 is similar to that for h_1 of Case 1, where error variance σ^2 has been assumed to be known; σ^2 is now replaced by its sample estimate, $\hat{\sigma}^2$.

(iii) It may be noted that the inversion formulae, through which the evaluation of the mixed moments is proposed to be done, are given for the order (a,b), satisfying $(a+b) \leq 6$. The doubt about the inadequacy of these formulae arises, naturally, when one finds that the mixed moments of orders (a,b), with $(a+b) > 6$ are also involved in the series in (3.3.33). This need not, however, be a matter of worry. The involvement of the term $\nu^{-(i+j+\frac{1}{2})}$ and the constants α_{ip}, β_{jq}, which decrease rapidly as

the series expand from one term to the next, assures that by the time the series would expand upto the term involving such a moment, the magnitude of the factor $[\alpha_{ip}\ \beta_{jq}\ \nu^{-(i+j+\frac{1}{2})}]$ would have already gone so low that the contribution of this term would be negligible. The same is true of the subsequent terms in the series. The proposed formulae are, therefore, adequate for the present purpose.

3.4 CASE 3 : UNKNOWN, HETEROGENEOUS VARIANCES OF ERRORS

We finally consider the situation in which the heterogeneity of variances, $\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2$, may seem to be a more reasonable assumption. Thus, in addition to the basic assumptions B(i) through B(v) we now assume that

C3(i) $\sigma_u^2 \neq \sigma_v^2$, $u, v = 1,2,\ldots,m$; $u \neq v$,

C3(ii) σ_u^2 is unknown, $u = 1,2,\ldots,m$.

With this set-up we derive an expression for $K_{u,v}(\underset{\sim}{\xi}_{n+1})$ to be used in the criterion function (2.4.5). Let Y_{n+1} be the random variable on which a discriminatory observation, y_{n+1}, is yet to be realized through a suitable setting of the input $\underset{\sim}{\xi}_{n+1}$. In the present state of ignorance we need the p.d.f. of Y_{n+1}, given n observations only. However, we start with the basic assumptions B(i) through B(v) only, ignoring C3(i) and C3(ii) for the time being. In that case, from the derivations in Section 3.2 we know that, given σ_u^2 $(= \lambda_u$ (say)) and n observations,

$\underset{\sim}{y}$, the p.d.f. of Y_{n+1}, under model u, can be expressed as [Refer equation (3.2.17)]

$$f^{(u)}_{n+1}(y_{n+1}/\lambda_u) = [2\pi\,\lambda_u(1+z_u)]^{-1/2} \exp\left[-\frac{1}{2}\left\{\frac{(y_{n+1} - \hat{y}^{(u)}_{n+1})^2}{\lambda_u(1+z_u)}\right\}\right], \tag{3.4.1}$$

where $\hat{y}^{(u)}_{n+1}$ is an estimated value of Y_{n+1}, using model u, z_u is given through the relation (3.2.14) and the partial derivatives involved therein are given by (3.2.4).

Nevertheless, in the present case, according to the assumption C3(ii), q_u is unknown. We, therefore, resort to the use of a noninformative prior for λ_u, i.e.,

$$g(\lambda_u) \propto \frac{1}{\lambda_u} .$$

Besides, we know that the sampling distribution of $(\nu_u\,\hat{\sigma}^2_u/\lambda_u)$ is given by

$$g(\hat{\sigma}^2_u/\lambda_u) = \frac{(\nu_u/2\lambda_u)^{\frac{\nu_u}{2}}}{\Gamma(\frac{u}{2})}\,(\hat{\sigma}^2_u)^{\frac{\nu_u}{2}-1} \exp\left(-\frac{\nu_u\,\hat{\sigma}^2_u}{2\lambda_u}\right),\ \lambda_u > 0, \tag{3.4.2}$$

where $\hat{\sigma}^2_u$ is an appropriate estimate of σ^2_u, obtained through model u alone and ν_u is the number of degrees of freedom attached with this estimate. After going through the same algebra as was used in proceeding from (3.3.1) to (3.3.9) in Section 3.3 (with σ^2 and $\hat{\sigma}^2$ replaced by σ^2_u and $\hat{\sigma}^2_u$, respectively) the required p.d.f. of Y_{n+1}, under model u, can be found to be of the form

$$f^{(u)}_{n+1}(y_{n+1}/\underset{\sim}{y}) = [B(\nu_u/2,1/2)\,\nu_u^{1/2}\,\hat{\zeta}_u]^{-1}\times$$

$$[1 + \frac{(y_{n+1} - \hat{y}^{(u)}_{n+1})}{\nu_u\,\hat{\zeta}^2_u}]^{-(\nu_u+1)/2}, \qquad (3.4.3)$$

where

$$\hat{\zeta}_u = \hat{\sigma}_u(1 + z_u)^{1/2}\ . \qquad (3.4.4)$$

It may be noted that the only prior information used in (3.4.3) comes from the previous n observations. Similarly, under model v($v \neq u$), the p.d.f. of Y_{n+1} is given by

$$f^{(v)}_{n+1}(y_{n+1}/\underset{\sim}{y}) = [B(\nu_v/2,1/2)\,\nu_v^{1/2}\hat{\zeta}_v]^{-1}\times$$

$$[1 + \frac{(y_{n+1} - \hat{y}^{(v)}_{n+1})^2}{\nu_v\,\hat{\zeta}^2_v}]^{-(\nu_v+1)/2}, \qquad (3.4.5)$$

with

$$\hat{\zeta}_v = \hat{\sigma}_v\,(1 + z_v)^{1/2}. \qquad (3.4.6)$$

In order to obtain an expression for measuring the dissimilarity between models u and v, the two densities $f^{(u)}_{n+1}$ and $f^{(v)}_{n+1}$ thus obtained are to be further utilized in the relation

$$h_3(f^{(u)}_{n+1}, f^{(v)}_{n+1}) = \int_{R^1} [f^{(u)}_{n+1}(y_{n+1}/\underset{\sim}{y})\ f^{(v)}_{n+1}(y_{n+1}/\underset{\sim}{y})]^{1/2}\,dy_{n+1}. \qquad (3.4.7)$$

In fact, as a result of substitution in (3.4.7) from (3.4.3) and (3.4.5) we get

$$h_3(f_{n+1}^{(u)}, f_{n+1}^{(v)}) =$$

$$[B(\nu_u/2,1/2)\ B(\nu_v/2,1/2)\ \nu_u^{1/2}\ \nu_v^{1/2}]^{-1/2}\ (\hat{\zeta}_u^2\ \hat{\zeta}_v^2)^{-1/4} \times$$

$$\int_{R^1} (1 + \frac{\tilde{t}_u}{\nu_u})^{-(\nu_u+1)/4}\ (1 + \frac{\tilde{t}_v}{\nu_v})^{-(\nu_v+1)/4}\ dy_{n+1}, \qquad (3.4.8)$$

where

$$\tilde{t}_i = \frac{(y_{n+1} - \hat{y}_{n+1}^{(i)})^2}{\hat{\zeta}_i^2}\ ,\ i = u,v\ . \qquad (3.4.9)$$

The integral in (3.4.8) is again intractable. However, it is of the same form as in (3.3.13) and can, therefore, be dealt with similarly. Proceeding on the same lines we can express each factor of the integrand in an alternative form. Thus, utilizing (3.3.22), the first factor can be written as

$$(1 + \frac{\tilde{t}_u}{\nu_u})^{-(\nu_u+1)/4} = \exp(-\tfrac{1}{4}\,\tilde{t}_u) \sum_{i=0}^{\infty} A_i(\tilde{t}_u)\ \nu_u^{-i}\ , \qquad (3.4.10)$$

where $A_i(\tilde{t}_u)$ is a polynomial in $\tilde{t}_u$ and is given by

$$A_i(\tilde{t}_u) = \sum_{p=1}^{i+1} \alpha_{ip}\ \tilde{t}_u^{2i-p+1}\ . \qquad (3.4.11)$$

Similarly, the second factor can be expressed as

$$(1 + \frac{\tilde{t}_v}{\nu_v})^{-(\nu_v+1)/4} = \exp(-\tfrac{1}{4}\,\tilde{t}_v) \sum_{j=0}^{\infty} B_j(\tilde{t}_v)\ \nu_v^{-j}, \qquad (3.4.12)$$

where

$$B_j(\tilde{t}_v) = \sum_{q=1}^{j+1} \beta_{jq}\, \tilde{t}_u^{2j-q+1} \quad . \tag{3.4.13}$$

The coefficients α_{ip}'s and β_{jq}'s appearing in (3.4.11) and (3.4.13) are given in Table 3.3.1. Multiplying the two factors from (3.4.10) and (3.4.12), we further have

$$(1 + \frac{\tilde{t}_u}{\nu_u})^{-(\nu_u+1)/4} \; (1 + \frac{\tilde{t}_v}{\nu_v})^{-(\nu_v+1)/4} =$$

$$\exp\{-\tfrac{1}{4}(\tilde{t}_u + \tilde{t}_v)\} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} Q_{ij}(\tilde{t}_u, \tilde{t}_v)\, \nu_u^{-i}\, \nu_v^{-j} \quad , \tag{3.4.14}$$

where Q_{ij} represents a bivariate polynomial in $\tilde{t}_u$ and $\tilde{t}_v$, obtained by multiplying $A_i(\tilde{t}_u)$ and $B_j(\tilde{t}_v)$ from (3.4.11) and (3.4.13), respectively, and is given by

$$Q_{ij}(\tilde{t}_u, \tilde{t}_v) = \sum_{p=1}^{i+1} \sum_{q=1}^{j+1} \alpha_{ip}\, \beta_{jq}\, \tilde{t}_u^{2i-p+1}\, \tilde{t}_v^{2j-q+1} \quad . \tag{3.4.15}$$

Now, using (3.4.9) the sum $(\tilde{t}_u + \tilde{t}_v)$ can be written as

$$\tilde{t}_u + \tilde{t}_v = \frac{(y_{n+1} - \hat{y}_{n+1}^{(u)})^2}{\hat{\zeta}_u^2} + \frac{(y_{n+1} - \hat{y}_{n+1}^{(v)})^2}{\hat{\zeta}_v^2} \quad .$$

But, through (A.1.11) and (A.1.12) of Corollary A.1 (Appendix A) we can, alternatively, express this sum as

$$\tilde{t}_u + \tilde{t}_v = \frac{(y_{n+1} - y_{n+1}^*)^2}{\tilde{\zeta}^2} + \frac{(\hat{y}_{n+1}^{(u)} - \hat{y}_{n+1}^{(v)})^2}{\hat{\zeta}_u^2 + \hat{\zeta}_v^2} \tag{3.4.16}$$

with $y^*_{n+1} = (\hat{\zeta}_v^2\, \hat{y}^{(u)}_{n+1} + \hat{\zeta}_u^2\, \hat{y}^{(v)}_{n+1})/(\hat{\zeta}_u^2 + \hat{\zeta}_v^2)$ and

$$\tilde{\zeta} = \{\frac{\hat{\zeta}_u^2 \hat{\zeta}_v^2}{(\hat{\zeta}_u^2 + \hat{\zeta}_v^2)}\}^{1/2} . \qquad (3.4.17)$$

The expression for $(\tilde{t}_u + \tilde{t}_v)$ from (3.4.16) when used in (3.4.14) further gives

$$(1 + \frac{\tilde{t}_u}{\nu_u})^{-(\nu_u+1)/4} (1 + \frac{\tilde{t}_v}{\nu_v})^{-(\nu_v+1)/4} =$$

$$\exp\ [-\tfrac{1}{4}\ \{\frac{(\hat{y}^{(u)}_{n+1} - \hat{y}^{(v)}_{n+1})^2}{\hat{\zeta}_u^2 + \hat{\zeta}_v^2}\}] \times$$

$$[\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} Q_{ij}(\tilde{t}_u, \tilde{t}_v)\, \nu_u^{-i}\, \nu_v^{-j}]\ \exp\ [-\tfrac{1}{4}\ \{\frac{(y_{n+1} - y^*_{n+1})^2}{\tilde{\zeta}^2}\}] .$$

(3.4.18)

Equation (3.4.18) presents an alternative expression for the integrand in (3.4.8). Subsequently, its use in (3.4.8) gives

$$h_3(f^{(u)}_{n+1}, f^{(v)}_{n+1}) =$$

$$[B(\nu_u/2, 1/2)\ B(\nu_v/2, 1/2)\ \nu_u^{1/2}\ \nu_v^{1/2} (\hat{\zeta}_u^2\ \hat{\zeta}_v^2)^{1/2}]^{-1/2} \times$$

$$\exp\ [-\tfrac{1}{4}\ \{\frac{(\hat{y}^{(u)}_{n+1} - \hat{y}^{(v)}_{n+1})^2}{\hat{\zeta}_u^2 + \hat{\zeta}_v^2}\}] \times$$

$$\int_{R^1} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} Q_{ij}(\tilde{t}_u, \tilde{t}_v)\, \nu_u^{-i}\, \nu_v^{-j}\ \exp[-\tfrac{1}{2}\{\frac{(y_{n+1} - y^*_{n+1})^2}{2\tilde{\zeta}^2}\}]\ dy_{n+1} .$$

(3.4.19)

We notice that the integral in (3.4.19) is of the same form as the integral in (3.3.31). It can, therefore, be likewise evaluated term by term. Thus, when Y_{n+1} in this case is considered to be distributed as $N_1(y^*_{n+1}, 2\tilde{\zeta}^2)$, as is suggested by the integrand in (3.4.19), and $\tilde{\mu}_{a,b}$ is used to denote $E(\tilde{t}_u^a \tilde{t}_v^b)$, the same argument enables us to write

$$h_3(f_{n+1}^{(u)}, f_{n+1}^{(v)}) =$$

$$[B(\nu_u/2,1/2)\ B(\nu_v/2,1/2)\ \nu_u^{1/2}\ \nu_v^{1/2}(\hat{\zeta}_u^2\ \hat{\zeta}_v^2)^{1/2}]^{-1/2} \times$$

$$\exp\ [-\tfrac{1}{4}\ \{\frac{(\hat{y}_{n+1}^{(u)} - \hat{y}_{n+1}^{(v)})^2}{\hat{\zeta}_u^2 + \hat{\zeta}_v^2}\}]\ (4\pi\ \tilde{\zeta}^2)^{1/2} \times$$

$$\sum_{i=0}^{\infty}\ \sum_{j=0}^{\infty}\ \sum_{p=1}^{i+1}\ \sum_{q=1}^{j+1}\ \alpha_{ip}\ \beta_{jq}\ \tilde{\mu}_{2i-p+1,2j-q+1}\ \nu_u^{-i}\ \nu_v^{-j}.$$

Finally, substituting for $\tilde{\zeta}$ from (3.4.17) and rearranging the terms we get

$$h_3(f_{n+1}^{(v)}, f_{n+1}^{(v)}) =$$

$$C_1^*\ [\frac{4\hat{\zeta}_u^2\hat{\zeta}_v^2}{(\hat{\zeta}_u^2 + \hat{\zeta}_v^2)^2}]^{1/4}\ \exp\ [-\tfrac{1}{4}\ \{\frac{(\hat{y}_{n+1}^{(u)} - \hat{y}_{n+1}^{(v)})}{(\hat{\zeta}_u^2 + \hat{\zeta}_v^2)}\}], \qquad (3.4.20)$$

where

$$C_1^* = [(2\pi)^{-1}\ B(\nu_u/2,1/2)\ B(\nu_v/2,1/2)]^{-1/2}$$

$$\sum_{i=0}^{\infty}\ \sum_{j=0}^{\infty}\ \sum_{p=1}^{i+1}\ \sum_{q=1}^{j+1}\ \alpha_{ip}\ \beta_{jq}\ \tilde{\mu}_{2i-p+1,2j-j+1}\ \nu_u^{-(i\ +\ \frac{1}{4})}\ \nu_v^{-(j\ +\ \frac{1}{4})} \qquad (3.4.21)$$

The required expression for $K_{u,v}(\underset{\sim}{\xi}_{n+1})$ to be used in the criterion function (2.4.5) for the present case would be the one obtained through the function $[1 - h_3(f_{n+1}^{(u)}, f_{n+1}^{(v)})]^{1/2}$.

As in the previous case, here, too, the mixed moments, $\tilde{\mu}_{ab}$, can be found through the joint cumulants $\tilde{k}_{ab}$ (say) by means of the inversion formulae of Cook (1951), listed in D.5 of Appendix D. So far as $\tilde{k}_{ab}$'s are concerned we use the same argument and from Lemma C.1 of Appendix C write the expression for the (a,b)th cumulant of $\tilde{t}_u$ and $\tilde{t}_v$ as

$$\tilde{k}_{ab} = 2^{a+b-1}\,(a+b-2)!\,[(a+b-1) + a(a-1)\,\tilde{\rho}\,d_u^2 + b(b-1)\,\tilde{\rho}\,d_v^2 + 2ab\,\tilde{\rho}\,d_u\,d_v]\,(\tilde{\rho}_u/\rho)^a\,(\tilde{\rho}_v/\rho)^b, \qquad (3.4.22)$$

where $\tilde{\rho}_s = (\hat{\zeta}_s^2)^{-1}$, $d_s = y_{n+1}^* - \hat{y}_{n+1}^{(s)}$, $s = u,v$, and $\rho = (\tilde{\rho}_u + \tilde{\rho}_v)/2$.

We close this section with a few remarks .

Remarks :

(i) So far as the divergence of the probability distributions for designing a new experiment is concerned, h_3 and hence $K_{u,v;n+1}$ considers dissimilarity of these distributions, in terms of the expectations and orientations, due to the observation being awaitted.

(ii) It may be noted that, except for the factor C_1^*, the form of h_3 is similar to that for h_2 of the previous case.

However, the estimate of the unknown σ_u^2 is no longer a pooled estimate.

(iii) Lastly, the inversion formulae, which seem to be falling short in this case, too, may similarly be confirmed to be adequate. This time, it is the factor $[\alpha_{ip}\,\beta_{jq}\,\nu_u^{-(i+\frac{1}{4})}\,\nu_v^{-(j+\frac{1}{4})}]$ which plays the role and renders the contribution of the moments $\tilde{\mu}_{a,b}, (a+b) > 6$ insignificant.

CHAPTER 4

DESIGN OF EXPERIMENTS FOR MODEL DISCRIMINATION IN MULTIRESPONSE SYSTEMS

4.1 BASIC ASSUMPTIONS

There are many fields in which the interest of the investigator often centres around two or more responses rather than a single response. Since there lies an advantage in considering all the responses together, it is important to formulate ways for designing experiments for model discrimination in a multiresponse system. In this chapter we plan to develop the design criteria under different types of assumptions which an investigator might think reasonable in a given situation.

In Chapter 2, we proposed that if the aim is to decide on an optimal experimental setting, the objective function

$$\phi(\underset{\sim}{\xi}_{n+1}) = \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} w_{u,v,n} K_{u,v}(\underset{\sim}{\xi}_{n+1})$$

may be maximized with respect to $\underset{\sim}{\xi}_{n+1}$ over the operability region. While the weights in this function have already been proposed to be obtained through the values of the discrimination index from the nth stage, the specification of the relevant form of $K_{u,v}(\underset{\sim}{\xi}_{n+1})$, under different types of conditions in a multiresponse system, is still pending. In this chapter we subject the

distance function, $K_{u,v}(\underset{\sim}{\xi}_{n+1})$, to different sets of assumptions. However, in order to develop the distribution theory necessary for deriving the particular forms, certain basic assumptions are required to be made.

We consider an r-response ($r \geq 2$) system and suppose that there are m r-equation models, $M^{(1)}, M^{(2)}, \ldots, M^{(m)}$, which are being discriminated. In model u, i.e., in the set of equations

$$Y_{ik} = \eta_i^{(u)}(\underset{\sim}{\xi}_k, \underset{\sim}{\theta}^{(u)}) + \varepsilon_{ik}^{(u)}, \quad i = 1,2,\ldots,r, \qquad (4.1.1)$$

we assume that

$B^+(i)$ $\quad E^{(u)}(\varepsilon_{ik}^{(u)}) = 0$,

$B^+(ii)$ $\quad E^{(u)}(\varepsilon_{ik}^{(u)}\varepsilon_{i\ell}^{(u)}) = 0$,

$B^+(iii)$ $\quad E^{(u)}(\varepsilon_{ik}^{(u)}\varepsilon_{jk}^{(u)}) = \sigma_{u;ij}$,

$B^+(iv)$ the error vector $\underset{\sim}{\varepsilon}_k = (\varepsilon_{k1}, \varepsilon_{k2}, \ldots, \varepsilon_{kr})$ is distributed as $N_r(\underset{\sim}{0}, \Sigma_u)$, where $\Sigma_u = \{\sigma_{u;ij}\}$;
$k, \ell = 1,2,\ldots,n,\ldots; \ell \neq k; i,j = 1,2,\ldots,r$,

$B^+(v)$ the model u, if nonlinear, can be approximated by a linear form in the parameter space;

the assumptions $B^+(i)$ through $B^+(v)$ being applicable to all the m models.

4.2 CASE 1: KNOWN, EQUAL COVARIANCE MATRICES

In a given situation, with all the above assumptions holding good, it may further be possible to assume that

$C^+(i)$ $\quad \Sigma_1 = \Sigma_2 = \ldots = \Sigma_m \ (= \Sigma)$,

$C^+(ii)$ the error covariance matrix Σ is known in advance.

We now seek the design criterion function when the models are nonlinear; the ones linear in parameters do not require any alteration in the results thus obtained. Let $\underset{\sim}{Y}_{n+1}$ be the random vector on which a discriminatory observation is being sought for. We first evaluate the p.d.f. of $\underset{\sim}{Y}_{n+1}$ under model u when the common covariance matrix, Σ, is given. With the basic assumptions holding good and the expectation $E^{(u)}(\underset{\sim}{Y}_{n+1})$ $[= \underset{\sim}{\mu}^{(u)}_{n+1}]$ as well as Σ being known, the p.d.f. of $\underset{\sim}{Y}_{n+1}$ is given by

$$f^{(u)}(\underset{\sim}{y}_{n+1}/\underset{\sim}{\mu}^{(u)}_{n+1}, \Sigma) =$$

$$[(2\pi)^r|\Sigma|]^{-1/2} \exp[-\tfrac{1}{2}\{(\underset{\sim}{y}_{n+1} - \underset{\sim}{\mu}^{(u)}_{n+1})' \ \Sigma^{-1}(\underset{\sim}{y}_{n+1} - \underset{\sim}{\mu}^{(u)}_{n+1})\}] . \qquad (4.2.1)$$

However, at the nth stage the expectation $\underset{\sim}{\mu}^{(u)}_{n+1}$ is not known. It is, therefore, the p.d.f. of $\underset{\sim}{Y}_{n+1}$, given Σ, which is required in the present set-up. In fact, the p.d.f. of $\underset{\sim}{Y}_{n+1}$, $f^{(u)}_{n+1}$ (say), in this case can be obtained through the relation

$$f^{(u)}_{n+1}(\underset{\sim}{y}_{n+1}/\Sigma) = \int_{R^r} f^{(u)}(\underset{\sim}{y}_{n+1}/\underset{\sim}{\mu}^{(u)}_{n+1}, \Sigma)\ g^{(u)}(\underset{\sim}{\mu}^{(u)}_{n+1}/\Sigma)\ d\underset{\sim}{\mu}^{(u)}_{n+1}. \qquad (4.2.2)$$

The first factor in the integrand in (4.2.2) is given in (4.2.1). In the following we seek an expression for the second factor. Now, under mondel u, we have

$$Y_{ik} = \eta^{(u)}_i(\underset{\sim}{\xi}_k, \underset{\sim}{\theta}^{(u)}) + \varepsilon^{(u)}_{ik} \quad . \qquad (4.2.3)$$

The assumption $B^{+}(v)$ permits the linearization of the nonlinear response function $\eta_i^{(u)}$ in the parameter space, R_θ (say). Therefore, if we decide to linearize the function about the m.l.e., $\hat{\underset{\sim}{\theta}}^{(u)}$, of $\underset{\sim}{\theta}^{(u)}$, then from (4.2.3) and by assumption $B^{+}(i)$ we can write

$$E^{(u)}(Y_{ik}) = \eta_i^{(u)}(\underset{\sim}{\xi}_k, \hat{\underset{\sim}{\theta}}^{(u)}) + \underset{\sim}{x}_{ik}^{(u)\prime}(\underset{\sim}{\theta}^{(u)} - \hat{\underset{\sim}{\theta}}^{(u)}), \quad i = 1,2,\ldots,r \qquad (4.2.4)$$

where

$$\underset{\sim}{x}_{ik}^{(u)} = (x_{ik1}^{(u)}, x_{ik2}^{(u)}, \ldots, x_{ikp_u}^{(u)})' \quad , \; i = 1,2,\ldots,r \; , \qquad (4.2.5)$$

and

$$x_{ikt}^{(u)} = \left[\frac{\partial \eta_i^{(u)}(\underset{\sim}{\xi}_k, \underset{\sim}{\theta}^{(u)})}{\partial \theta_t^{(u)}}\right]_{\underset{\sim}{\theta}^{(u)} = \hat{\underset{\sim}{\theta}}^{(u)}} \quad , \qquad (4.2.6)$$

$$t = 1,2,\ldots,p_u \; .$$

The relation (4.2.4) can also be written as

$$E^{(u)}(\underset{\sim}{Y}_k) = \underset{\sim}{\eta}^{(u)}(\underset{\sim}{\xi}_k, \hat{\underset{\sim}{\theta}}^{(u)}) + X_k^{(u)}(\underset{\sim}{\theta}^{(u)} - \hat{\underset{\sim}{\theta}}^{(u)}), \qquad (4.2.7)$$

where $\underset{\sim}{Y}_k$ is an r-vector and

$$X_k^{(u)} = (\underset{\sim}{x}_{1k}^{(u)}, \underset{\sim}{x}_{2k}^{(u)}, \ldots, \underset{\sim}{x}_{rk}^{(u)})' \; . \qquad (4.2.8)$$

$X_k^{(u)}$ is an $r \times p_u$ matrix of partial derivatives, $x_{ikt}^{(u)}$, given by (4.2.6). In particular, from (4.2.7) we have

$$\underset{\sim}{\mu}_{n+1}^{(u)} = \underset{\sim}{\eta}^{(u)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}^{(u)}) + X_{n+1}^{(u)}(\underset{\sim}{\theta}^{(u)} - \hat{\underset{\sim}{\theta}}^{(u)}), \qquad (4.2.9)$$

i.e.,

$$(\underset{\sim}{\mu}_{n+1}^{(u)} - \hat{\underset{\sim}{Y}}_{n+1}^{(u)}) = X_{n+1}^{(u)}(\underset{\sim}{\theta}^{(u)} - \hat{\underset{\sim}{\theta}}^{(u)}) \quad , \qquad (4.2.10)$$

where

$$\hat{\underset{\sim}{y}}_{n+1}^{(u)} = \underset{\sim}{\eta}^{(u)}(\underset{\sim}{\xi}_{n+1}, \hat{\underset{\sim}{\theta}}^{(u)}) . \qquad (4.2.11)$$

This suggests that the distribution of $(\underset{\sim}{\mu}_{n+1}^{(u)} - \hat{\underset{\sim}{Y}}_{n+1}^{(u)})$ is identical with that of $\underset{n+1}{X}^{(u)} (\underset{\sim}{\theta}^{(u)} - \hat{\underset{\sim}{\theta}}^{(u)})$. In the following we adopt the Bayesian approach and seek the posterior distribution of $(\underset{\sim}{\theta}^{(u)} - \hat{\underset{\sim}{\theta}}^{(u)})$.

Now, the likelihood function of n sets of data, $\underset{\approx}{y}$, under model u, is given by

$$L(\underset{\sim}{\theta}^{(u)}/\underset{\approx}{y}, \Sigma) = [(2\pi)^r |\Sigma|]^{-n/2} \exp\left[-\frac{1}{2} \sum_{k=1}^{n} \sum_{j=1}^{r} \sum_{j=1}^{r} \sigma_u^{ij} v_{ijk}^{(u)}\right], \qquad (4.2.12)$$

where $\underset{\approx}{y}$, an $n \times r$ matrix, represents the whole set of data, σ_u^{ij} denotes the (ij)th element of Σ_u^{-1}, the inverse of Σ_u, and

$$v_{ijk}^{(u)} = \{y_{ik} - \eta_i^{(u)}(\underset{\sim}{\xi}_k, \hat{\underset{\sim}{\theta}}^{(u)})\}\{y_{jk} - \eta_j^{(u)}(\underset{\sim}{\xi}_k, \hat{\underset{\sim}{\theta}}^{(u)})\} .$$

If we let $e_{ik}^{(u)} = \{y_{ik} - \eta_i^{(u)}(\underset{\sim}{\xi}_k, \hat{\underset{\sim}{\theta}}^{(u)})\}$, $i = 1,2,\ldots,r$, then using the linear form (4.2.4) of the response function $\eta_i^{(u)}$ we can write

$$v_{ijk}^{(u)} = \{e_{ik}^{(u)} - \underset{\sim}{x}_{ik}^{(u)'}(\underset{\sim}{\theta}^{(u)} - \hat{\underset{\sim}{\theta}}^{(u)})\}\{e_{jk}^{(u)} - \underset{\sim}{x}_{jk}^{(u)'}(\underset{\sim}{\theta}^{(u)} - \hat{\underset{\sim}{\theta}}^{(u)})\} .$$

Further, using the fact that $\hat{\underset{\sim}{\theta}}^{(u)}$ is an m.l.e. of $\underset{\sim}{\theta}^{(u)}$, we have $e_{ik}^{(u)}\, \underset{\sim}{x}_{jk}^{(u)} = \underset{\sim}{0}$. Consequently,

$$v_{ijk}^{(u)} = e_{ik}^{(u)}\, e_{jk}^{(u)} + (\underset{\sim}{\theta}^{(u)} - \hat{\underset{\sim}{\theta}}^{(u)})'\, \underset{\sim}{x}_{ik}^{(u)}\, \underset{\sim}{x}_{jk}^{(u)'}(\underset{\sim}{\theta}^{(u)} - \hat{\underset{\sim}{\theta}}^{(u)}) .$$

The likelihood function L in (4.2.12) thus becomes

$$L(\underset{\sim}{\theta}^{(u)}/\underset{\approx}{y},\Sigma) = [(2\pi)^r|\Sigma|]^{-n/2} \exp[-\tfrac{1}{2}\sum_{k=1}^{n}\sum_{i=1}^{r}\sum_{j=1}^{r}\sigma_u^{ij}\, e_{ik}^{(u)}\, e_{jk}^{(u)}] \times$$

$$[\exp\{-\tfrac{1}{2}(\underset{\sim}{\theta}^{(u)} - \underset{\sim}{\hat{\theta}}^{(u)})'\, M_u\, (\underset{\sim}{\theta}^{(u)} - \underset{\sim}{\hat{\theta}}^{(u)})\}], \tag{4.2.13}$$

where

$$M_u = \sum_{k=1}^{n}\sum_{i=1}^{r}\sum_{j=1}^{r}\sigma^{ij}\, \underset{\sim}{x}_{ik}^{(u)}\, \underset{\sim}{x}_{jk}^{(u)'} \,. \tag{4.2.14}$$

Now, if we assume a locally uniform prior distribution for $\underset{\sim}{\theta}^{(u)}$, then the posterior probability density of $\underset{\sim}{\theta}^{(u)}$ is given by

$$g(\underset{\sim}{\theta}^{(u)}/\underset{\approx}{y},\Sigma) = \frac{L(\underset{\sim}{\theta}^{(u)}/\underset{\approx}{y},\Sigma)}{\int_{R_\theta} L(\underset{\sim}{\theta}^{(u)}/\underset{\approx}{y},\Sigma)\, d\underset{\sim}{\theta}^{(u)}} \,. \tag{4.2.15}$$

Using, therefore, the likelihood (4.2.13) in (4.2.15) we get

$$g(\underset{\sim}{\theta}^{(u)}/\underset{\approx}{y},\Sigma) = [(2\pi)^{-p_u}|M_u|]^{1/2} \exp[-\tfrac{1}{2}\{(\underset{\sim}{\theta}^{(u)} - \underset{\sim}{\hat{\theta}}^{(u)})' M_u\, (\underset{\sim}{\theta}^{(u)} - \underset{\sim}{\hat{\theta}}^{(u)})\}], \tag{4.2.16}$$

where M_u is given by (4.2.14).

This shows that the posterior probability distribution of $(\underset{\sim}{\theta}^{(u)} - \underset{\sim}{\hat{\theta}}^{(u)})$ is a p_u-variate normal istribution with mean vector $\underset{\sim}{0}$ and covariance matrix M_u^{-1}. Further, the r-vector $X_{n+1}^{(u)}\,(\underset{\sim}{\theta}^{(u)} - \underset{\sim}{\hat{\theta}}^{(u)})$ being a linear combination of the normal variates is an r-variate normal random vector, distributed about $\underset{\sim}{0}$ with covariance matrix $[X_{n+1}^{(u)}\, M_u^{-1}\, X_{n+1}^{(u)'}] (= C_u$ (say)). Thus the

relation (4.2.10) leads us to the conclusion that the vector $(\underset{\sim}{\mu}_{n+1}^{(u)} - \underset{\sim}{\hat{y}}_{n+1}^{(u)})$ has an r-variate normal distribution : $N_r(\underset{\sim}{0}, C_u)$. Accordingly, the second factor of the integrand in (4.2.2) can be written as

$$g^{(u)}(\underset{\sim}{\mu}_{n+1}^{(u)}/\Sigma) =$$

$$[(2\pi)^r |C_u|]^{-1/2} \exp\left[-\tfrac{1}{2}\{(\underset{\sim}{\mu}_{n+1}^{(u)} - \underset{\sim}{\hat{y}}_{n+1}^{(u)})' \, C_u^{-1} (\underset{\sim}{\mu}_{n+1}^{(u)} - \underset{\sim}{\hat{y}}_{n+1}^{(u)})\}\right]. \tag{4.2.17}$$

Substituting for $f^{(u)}$ and $g^{(u)}$ in (4.2.2) from (4.2.1) and (4.2.17), respectively, we get

$$f_{n+1}^{(u)}(\underset{\sim}{y}_{n+1}/\Sigma) = (2\pi)^r (|C_u|)^{-1/2} \int_{R^r} \exp\{-\tfrac{1}{2} Q_1(\underset{\sim}{\mu}_{n+1}^{(u)})\} \, d\underset{\sim}{\mu}_{n+1}^{(u)}, \tag{4.2.18}$$

where

$$Q_1(\underset{\sim}{\mu}_{n+1}^{(u)}) = \{(\underset{\sim}{\mu}_{n+1}^{(u)} - \underset{\sim}{y}_{n+1})' \, \Sigma^{-1} \, (\underset{\sim}{\mu}_{n+1}^{(u)} - \underset{\sim}{y}_{n+1})\} +$$

$$\{(\underset{\sim}{\mu}_{n+1}^{(u)} - \underset{\sim}{\hat{y}}_{n+1}^{(u)})' \, C_u^{-1} (\underset{\sim}{\mu}_{n+1}^{(u)} - \underset{\sim}{\hat{y}}_{n+1}^{(u)})\} \, . \tag{4.2.19}$$

Now, using equations (A.1.1) through (A.1.3) of Lemma A.1 (Appendix A), the function $Q_1(\underset{\sim}{\mu}_{n+1}^{(u)})$ can alternatively be expressed as

$$Q_1(\underset{\sim}{\mu}_{n+1}^{(u)}) = \{(\underset{\sim}{\mu}_{n+1}^{(u)} - \underset{\sim}{\bar{\mu}}_{n+1})' \, (\Sigma^{-1} + C_u^{-1}) \, (\underset{\sim}{\mu}_{n+1}^{(u)} - \underset{\sim}{\bar{\mu}}_{n+1})\} +$$

$$\{(\underset{\sim}{y}_{n+1} - \underset{\sim}{\hat{y}}_{n+1}^{(u)})' \, (\Sigma + C_u)^{-1} \, (\underset{\sim}{y}_{n+1} - \underset{\sim}{\hat{y}}_{n+1}^{(u)})\} \, , \tag{4.2.20}$$

where $\bar{\bar{\mu}}_{n+1}$ is given by

$$\bar{\bar{\mu}}_{n+1} = (\Sigma^{-1} + C_u^{-1})^{-1} (\Sigma^{-1} y_{n+1} + C_u^{-1} \hat{y}_{n+1}^{(u)}) \quad . \qquad (4.2.21)$$

Making use of this form of $Q_1(\mu_{n+1}^{(u)})$ in (4.2.18) we get

$$f_{n+1}^{(u)}(y_{n+1}/\Sigma) =$$

$$[\{(2\pi)^{-r} |\Sigma^{-1}| |C_u^{-1}|\}^{1/2}] \times$$

$$\exp [-\tfrac{1}{2}\{(y_{n+1} - \hat{y}_{n+1}^{(u)})' (\Sigma + C_u)^{-1} (y_{n+1} - \hat{y}_{n+1}^{(u)})\}] \times$$

$$(2\pi)^{-r/2} \int_{R^r} \exp [-\tfrac{1}{2}\{(\mu_{n+1}^{(u)} - \bar{\bar{\mu}}_{n+1})' (\Sigma^{-1}+C_u^{-1})(\mu_{n+1}^{(u)} - \bar{\bar{\mu}}_{n+1})\}] d\mu_{n+1}^{(u)},$$

i.e.,

$$f_{n+1}^{(u)}(y_{n+1}/\Sigma) = [(2\pi)^{-r} |\Sigma^{-1}| |C_u^{-1}| |\Sigma^{-1} + C_u^{-1}|]^{1/2} \times$$

$$\exp [-\tfrac{1}{2}\{(y_{n+1} - \hat{y}_{n+1}^{(u)})' (\Sigma+C_u)^{-1}(y_{n+1} - \hat{y}_{n+1}^{(u)})\}] . \qquad (4.2.22)$$

Since Σ^{-1} and C_u^{-1} exist we can use the identity,

$$\Sigma^{-1} + C_u^{-1} = \Sigma^{-1} (\Sigma + C_u) C_u^{-1}$$

in (4.2.22) and conclude that given Σ and n sets of observations $y_1, y_2, \ldots, y_n$, the p.d.f. of Y_{n+1} under model u is given by

$$f_{n+1}^{(u)}(y_{n+1}/\Sigma) = [(2\pi)^r |Z_u|]^{-1/2} \times$$

$$\exp[-\tfrac{1}{2}\{(y_{n+1} - \hat{y}_{n+1}^{(u)})' Z_u^{-1}(Y_{n+1} - \hat{y}_{n+1}^{(u)})\}] , \qquad (4.2.23)$$

where

$$Z_u = \Sigma + C_u \quad . \tag{4.2.24}$$

Similarly, given Σ and n sets of observations, the p.d.f. of $\underset{\sim}{Y}_{n+1}$ under model v would be

$$f_{n+1}^{(v)}(\underset{\sim}{y}_{n+1}/\Sigma) = [(2\pi)^r |Z_v|]^{-1/2} \times \exp\left[-\tfrac{1}{2}\{(\underset{\sim}{y}_{n+1} - \hat{\underset{\sim}{y}}_{n+1}^{(v)})' Z_v^{-1} (\underset{\sim}{y}_{n+1} - \hat{\underset{\sim}{y}}_{n+1}^{(v)})\}\right] \tag{4.2.25}$$

with

$$Z_v = \Sigma + C_v \quad . \tag{4.2.26}$$

Consider now the function

$$h_1(f_{n+1}^{(u)}, f_{n+1}^{(v)}) = \int_{R^r} (f_{n+1}^{(u)} f_{n+1}^{(v)})^{1/2} \, d\underset{\sim}{y}_{n+1} \quad . \tag{4.2.27}$$

Substituting for $f_{n+1}^{(u)}$ and $f_{n+1}^{(v)}$ in (4.2.27) from (4.2.23) and (4.2.25) we get

$$h_1(f_{n+1}^{(u)}, f_{n+1}^{(v)}) = [(2\pi)^{2r} |Z_u| \, |Z_v|]^{-1/4} \int_{R^r} \exp\{-\tfrac{1}{4} Q_2(\underset{\sim}{y}_{n+1})\} d\underset{\sim}{y}_{n+1} \, , \tag{4.2.28}$$

where

$$Q_2(\underset{\sim}{y}_{n+1}) = \{(\underset{\sim}{y}_{n+1} - \hat{\underset{\sim}{y}}_{n+1}^{(u)})' \, Z_u^{-1} (\underset{\sim}{y}_{n+1} - \hat{\underset{\sim}{y}}_{n+1}^{(u)})\} + \{(\underset{\sim}{y}_{n+1} - \hat{\underset{\sim}{y}}_{n+1}^{(v)})' Z_v^{-1} (\underset{\sim}{y}_{n+1} - \hat{\underset{\sim}{y}}_{n+1}^{(v)})\} \quad .$$

Using (A.1.1) through (A.1.3) of Lemma A.1 (Appendix A) we can write $Q_2(\underset{\sim}{y}_{n+1})$ as

$$Q_2(\underset{\sim}{y}_{n+1}) = \{(\underset{\sim}{y}_{n+1} - \underset{\sim}{y}^*_{n+1})' \, Z^*(\underset{\sim}{y}_{n+1} - \underset{\sim}{y}^*_{n+1})\} +$$

$$\{(\hat{\underset{\sim}{y}}^{(u)}_{n+1} - \hat{\underset{\sim}{y}}^{(v)}_{n+1})' (Z_u + Z_u)^{-1} (\hat{\underset{\sim}{y}}^{(u)}_{n+1} - \hat{\underset{\sim}{y}}^{(v)}_{n+1})\},$$

where

$$\underset{\sim}{y}^*_{n+1} = (Z_u + Z_v)^{-1} (Z_v \, \hat{\underset{\sim}{y}}^{(u)}_{n+1} + Z_u \, \hat{\underset{\sim}{y}}^{(v)}_{n+1}) \quad \text{and}$$

$$Z^* = Z_u^{-1} (Z_u + Z_v) \, Z_v^{-1} . \qquad (4.2.29)$$

The expression for $Q_2(\underset{\sim}{y}_{n+1})$, when used in (4.2.28), gives

$$h_1(f^{(u)}_{n+1}, f^{(v)}_{n+1}) =$$

$$[(2\pi)^{2r} \, |Z_u| \, |Z_v|]^{-1/4} \times$$

$$\exp\left[-\tfrac{1}{4} \{(\hat{\underset{\sim}{y}}^{(u)}_{n+1} - \hat{\underset{\sim}{y}}^{(v)}_{n+1})' \, (Z_u + Z_v)^{-1} \, (\hat{\underset{\sim}{y}}^{(u)}_{n+1} - \hat{\underset{\sim}{y}}^{(v)}_{n+1})\}\right] \times$$

$$\int_{R^r} \exp\left[-\tfrac{1}{4} \{(\underset{\sim}{y}_{n+1} - \underset{\sim}{y}^*_{n+1})' \, Z^*(\underset{\sim}{y}_{n+1} - \underset{\sim}{y}^*_{n+1})\}\right] d\underset{\sim}{y}_{n+1}$$

$$= [|Z_u| \, |Z_v| 4 |Z^*|^2]^{-1/4} \times$$

$$\exp\left[-\tfrac{1}{4} \{(\hat{\underset{\sim}{y}}^{(u)}_{n+1} - \hat{\underset{\sim}{y}}^{(v)}_{n+1})' \, (Z_u + Z_v)^{-1} \, (\hat{\underset{\sim}{y}}^{(u)}_{n+1} - \hat{\underset{\sim}{y}}^{(v)}_{n+1})\}\right] .$$

Substituting for Z^* from (4.2.29) and rearranging the terms we get

$$h_1(f^{(u)}_{n+1}, f^{(v)}_{n+1}) =$$

$$\left[\frac{4|Z_u Z_v|}{|Z_u + Z_v|^2}\right]^{1/4} \exp\left[-\tfrac{1}{4}\{(\hat{\underset{\sim}{y}}^{(u)}_{n+1} - \hat{\underset{\sim}{y}}^{(v)}_{n+1})' \, (Z_u + Z_v)^{-1} (\hat{\underset{\sim}{y}}^{(u)}_{n+1} - \hat{\underset{\sim}{y}}^{(v)}_{n+1})\}\right] . \qquad (4.2.30)$$

Finally, utilizing the expression for h_1 from (4.2.30) in the function

$$K_{u,v}(\underset{\sim}{\xi}_{n+1}) = [1 - h_1(f_{n+1}^{(u)}, f_{n+1}^{(v)})]^{1/2}$$

we can obtain the required distance function to be used in the criterion function $\phi(\underset{\sim}{\xi}_{n+1})$ for designing the (n+1)th experiment, when n sets of observations are available, Σ, the common covariance matrix, is known, and the models are linear or have been linearized appropriately.

4.3 CASE 2: UNKNOWN, EQUAL ERROR COVARIANCE MATRICES

The design criterion developed in the last section cannot be used if Σ, the common covariance matrix, is not known, even though all the basic assumptions hold good and the requirement of equality of covariance matrices is met. In fact, there are situations in which, even if the investigator has reasons to believe that the errors in different models have same covariance structure the knowledge about the common covariance matrix is lacking. In this section, we, therefore, develop the design criterion through which these types of situations could be suitably handled. Thus, whereas we retain the assumption $B^+(i)$ through $B^+(v)$ we assume as well that

$C^+2(i)$ $\quad \Sigma_1 = \Sigma_2 = \ldots = \Sigma_m \ (= \Sigma)$,

$C^+2(ii)$ $\quad \Sigma$, the common covariance matrix, is unknown.

Let $\underset{\sim}{Y}_{n+1}$ be the random vector on which a discriminatory observation is awaitted. In order to plan an experiment $\underset{\sim}{\xi}_{n+1}$ which would yield such a response under the given set of assumptions; namely, $B^+(i)$ through $B^+(v)$, $C^+2(i)$, and $C^+2(ii)$, we must, first, find the probability distribution of $\underset{\sim}{Y}_{n+1}$ when Σ is unknown. With all the assumptions of the last section holding good, we know that, under model u, $\underset{\sim}{Y}_{n+1}$ is distributed normally with mean vector $\hat{\underset{\sim}{y}}^{(u)}_{n+1}$ and covariance matrix Z_u, with Z_u given by

$$Z_u = [\Sigma + X^{(u)}_{n+1} (\sum_{k=1}^{n} \sum_{i=1}^{r} \sum_{j=1}^{r} \sigma^{ij} x^{(u)}_{ik} x^{(u)'}_{jk})^{-1} x^{(u)'}_{n+1}]. \qquad (4.3.1)$$

Presently, we are considering a situation in which the covariance matrix Σ and hence Σ^{-1} $(= \{\sigma^{ij}\})$ is not known in advance. If we still plan to make an allowance for the variances and covariances of errors in the design of experiments we ought to fall back upon the sample and use their estimates so as to obtain an estimate of Z_u. We propose to use a plug-in-estimate, namely,

$$\hat{Z}_u = [\hat{\Sigma} + X^{(u)}_{n+1} (\sum_{k=1}^{n} \sum_{i=1}^{r} \sum_{j=1}^{r} \hat{\sigma}^{ij} x^{(u)}_{ik} x^{(u)'}_{jk})^{-1} x^{(u)'}_{n+1}], \qquad (4.3.2)$$

where, because of the assumption $C^+2(i)$, the estimate $\hat{\Sigma}$ is obtained by pooling the estimates, $\hat{\Sigma}_u$, obtained through the m rival models, i.e.,

$$\hat{\Sigma} = \frac{\sum_{u=1}^{m} \nu_u \hat{\Sigma}_u}{\nu}, \qquad (4.3.3)$$

where $\nu = \sum_{u=1}^{m} \nu_u$ is the number of degrees of freedom associated with the pooled estimate $\hat{\Sigma}$.

It may be noted that because of this new arrangement, $\underset{\sim}{Y}_{n+1}$ no longer obeys a normal probability law. As in the univariate case, we replace the r-variate normal distribution of $\underset{\sim}{Y}_{n+1}$ by an r-variate t distribution : $t_r\ (\hat{\underset{\sim}{y}}^{(u)}_{n+1},\ \hat{Z}_u, \nu)$, where $\hat{Z}_u$ is an estimate of Z_u given by (4.3.2) and ν is the number of degrees of freedom associated with this estimation. Thus, given n sets of observations, the p.d.f. of $\underset{\sim}{Y}_{n+1}$, under model u, can be specified as

$$f^{(u)}_{n+1}(\underset{\sim}{y}_{n+1}/\underset{\approx}{y}) = \left[\frac{\Gamma(\frac{\nu+r}{2})}{\{\Gamma(\frac{1}{2})\}^r\ \Gamma(\frac{\nu}{2})\ \nu^{r/2}}\right] |\hat{Z}_u|^{-1/2} \times$$

$$\left[1+(\underset{\sim}{y}_{n+1} - \hat{\underset{\sim}{y}}^{(u)}_{n+1})'\ (\nu\ \hat{Z}_u)^{-1}\ (\underset{\sim}{y}_{n+1} - \hat{\underset{\sim}{y}}^{(u)}_{n+1})\right]^{-(\nu+r)/2}. \qquad (4.3.4)$$

Similarly, given n sets of observations, the p.d.f. of $\underset{\sim}{Y}_{n+1}$ under model v, can be written as

$$f^{(v)}_{n+1}(\underset{\sim}{y}_{n+1}/\underset{\approx}{y}) = \left[\frac{\Gamma(\frac{\nu+r}{2})}{\{\Gamma(\frac{1}{2})\}^r\ \Gamma(\frac{\nu}{2})\ \nu^{r/2}}\right] |\hat{Z}_v|^{-1/2} \times$$

$$\left[1+(\underset{\sim}{y}_{n+1} - \hat{\underset{\sim}{y}}^{(v)}_{n+1})'\ (\nu\ \hat{Z}_v)^{-1}\ (\underset{\sim}{y}_{n+1} - \hat{\underset{\sim}{y}}^{(v)}_{n+1})\right]^{-(\nu+r)/2}, \qquad (4.3.5)$$

where

$$\hat{Z}_v = [\hat{\Sigma} + X_{n+1}^{(v)} \left(\sum_{k=1}^{n} \sum_{i=1}^{r} \sum_{j=1}^{r} \hat{\sigma}^{ij} X_{ik}^{(v)} X_{jk}^{(v)'}\right)^{-1} X_{n+1}^{(v)'}] \tag{4.3.6}$$

with ν degrees of freedom.

Having decided on the alternative probability distributions of $\underset{\sim}{Y}_{n+1}$ under models u and v with the restriction that Σ, the common covariance matrix, is unknown, we now seek the divergence

$$h_2(f_{n+1}^{(u)}, f_{n+1}^{(u)}) = \int_{R^r} [f_{n+1}^{(u)}(\underset{\sim}{y}_{n+1}/\underset{\approx}{y})\ f_{n+1}^{(v)}(\underset{\sim}{y}_{n+1}/\underset{\approx}{y})]^{1/2}\ d\underset{\sim}{y}_{n+1} \tag{4.3.7}$$

Substituting for $f_{n+1}^{(u)}$ and $f_{n+1}^{(v)}$ in (4.3.7) from (4.3.4) and (4.3.5) we get

$$h_2(f_{n+1}^{(u)}, f_{n+1}^{(v)}) = C_1\ |\hat{Z}_u\ \hat{Z}_v|^{-1/4} \int_{R^r} Q_3(\underset{\sim}{y}_{n+1})\ d\underset{\sim}{y}_{n+1}, \tag{4.3.8}$$

where

$$C_1 = \left[\frac{\Gamma(\frac{\nu+r}{2})}{\{\Gamma(\frac{1}{2})\}^r\ \Gamma(\frac{\nu}{2})\ \nu^{r/2}}\right], \tag{4.3.9}$$

$$Q_3(\underset{\sim}{y}_{n+1}) = [1 + T_u\ \nu^{-1}]^{-(\nu+r)/4}\ [1 + T_v\ \nu^{-1}]^{-(\nu+r)/4}, \tag{4.3.10}$$

and

$$T_i = (\underset{\sim}{y}_{n+1} - \underset{\sim}{\hat{y}}_{n+1}^{(i)})'\ \hat{Z}_i^{-1}\ (\underset{\sim}{y}_{n+1} - \underset{\sim}{\hat{y}}_{n+1}^{(i)}),\ i = u,v\ . \tag{4.3.11}$$

The integrand Q_3 in (4.3.8) can be simplified to a tractable form. To that purpose we write

$$[1 + T_u \nu^{-1}]^{-(\nu+r)/4} = \exp \{- (\tfrac{\nu+r}{4}) \log_e (1 + T_u \nu^{-1})\}. \tag{4.3.12}$$

Now, if we use the expansion of $\log_e (1 + T_u \nu^{-1})$ in powers of $T_u \nu^{-1}$ on the right hand side of (4.3.12) and arrange the terms in powers of ν^{-1}, we can write

$$[1 + T_u \nu^{-1}]^{-(\nu+r)/4} = \exp (-\tfrac{1}{4} T_u) \exp (\sum_{i=1}^{\infty} \tau_i \nu^{-i}), \tag{4.3.13}$$

where

$$\tau_i = \frac{(-1)^i}{4i(i+1)} \{(i+1)r T_u^i - i T_u^{(i+1)}\}. \tag{4.3.14}$$

(Refer D.6 of Appendix D for details of the algebra used). Further, the second factor on the right hand side of (4.3.13) can be written as

$$\exp (\sum_{i=1}^{\infty} \tau_i \nu^{-i}) = \prod_{i=1}^{\infty} \exp (\tau_i \nu^{-i}),$$

so that on using the expansions of the factors $\exp (\tau_i \nu^{-i})$ in the infinite product and cclllecting coefficients of like powers of ν^{-1} we obtain

$$\exp (\sum_{i=1}^{\infty} \tau_i \nu^{-i}) = \sum_{i=0}^{\infty} a_i \nu^{-i}, \tag{4.3.15}$$

where

$$a_0 = 1,$$

$$a_1 = \tau_1,$$

$$a_2 = \frac{\tau_1^2}{2} + \tau_2 ,$$

$$a_3 = \frac{\tau_1^3}{6} + \tau_1\tau_2 + \tau_3 ,$$

$$a_4 = \frac{\tau_1^4}{24} + \frac{\tau_1^2\tau_2}{2} + \frac{\tau_2^2}{2} + \tau_1\tau_3 + \tau_4 ,$$

$$a_5 = \frac{\tau_1^5}{120} + \frac{\tau_1^3\tau_2}{6} + \frac{\tau_1\tau_2^2}{2} + \frac{\tau_1^2\tau_3}{2} + \tau_1\tau_4 + \tau_2\tau_3 + \tau_5 ,$$

$$a_6 = \frac{\tau_1^6}{720} + \frac{\tau_1^4\tau_2}{24} + \frac{\tau_1^2\tau_4}{2} + \frac{\tau_1^2\tau_2^2}{4} + \frac{\tau_1^3\tau_3}{6} + \frac{\tau_2^3}{6} + \tau_1\tau_2\tau_3 + \frac{\tau_3^2}{2} + \tau_1\tau_5 + \tau_2\tau_4 + \tau_6 ,$$

etc.

In order to express the series on the right hand side of (4.3.15) in terms of T_u we use the relation (4.3.14) and get

$$\exp\left(\sum_{i=1}^{\infty} \tau_i \nu^{-i}\right) = \sum_{i=0}^{\infty} A_i(T_u)\, \nu^{-i}, \qquad (4.3.16)$$

where

$$A_i(T_u) = \sum_{k=1}^{i+1} \sum_{s=1}^{k} \alpha_{iks}\, r^{k-s}\, T_u^{2i-k+1} \qquad (4.3.17)$$

and α_{iks}'s in (4.2.15) are given in Table 4.3.1. Hence from (4.3.13) we get

$$[1 + T_u \nu^{-1}]^{-(\nu+r)/4} = \exp\left(-\frac{1}{4} T_u\right) \sum_{i=0}^{\infty} A_i(T_u)\, \nu^{-i} . \qquad (4.3.18)$$

Similarly, the second factor of the integrand Q_3 in (4.3.8) can be expressed as

Table 4.3.1
Values of the coefficients α_{iks} and $\beta_{j\ell t}$

k,s;ℓ,t	i,j 0	1	2	3	4	5	6
1 1	1.000	0.125	0.016	3.255E-4	1.017E-5	2.543E-7	5.298E-9
2 1		0.250	0.063	1.953E-3	8.138E-5	2.543E-6	6.358E-8
2		0.000	0.083	0.010	6.510E-4	2.713E-5	8.477E-7
3 1			0.063	3.906E-3	2.441E-4	1.017E-5	3.179E-7
2			0.125	0.036	3.906E-3	1.420E-4	8.053E-6
3			0.000	0.063	0.011	9.223E-4	4.747E-5
4 1				2.604E-3	3.255E-4	2.034E-5	8.477E-7
2				0.031	6.510E-3	5.697E-4	3.052E-5
3				0.083	0.036	4.557E-3	3.391E-4
4				0.000	0.050	0.011	1.138E-3
5 1					1.628E-4	2.034E-5	1.272E-6
2					3.906E-3	7.053E-4	5.764E-5
3					0.029	8.138E-3	9.020E-4
4					0.062	0.035	5.631E-3
5					0.000	0.042	0.011
6 1						8.138E-6	1.017E-6
2						3.255E-4	5.425E-5
3						4.557E-3	1.058E-3
4						0.026	9.116E-3
5						0.050	0.003
6						0.000	0.036
7 1							3.391E-7
2							2.034E-5
3							4.611E-4
4							4.883E-3
5							0.024
6							0.042
7							0.000

E-k stands for $\times 10^{-k}$

$$[1 + T_v \tau^{-1}]^{-(\nu+r)/4} = \exp\left(-\frac{1}{4} T_v\right) \sum_{j=0}^{\infty} B_j(T_v)\, \nu^{-j}, \tag{4.3.19}$$

where

$$B_j(T_v) = \sum_{\ell=1}^{j+1} \sum_{t=1}^{\ell} \beta_{j\ell t}\, r^{\ell-t}\, T_v^{2j-\ell+1} \tag{4.3.20}$$

and $\beta_{j\ell t}$'s are given in Table 4.3.1.

As a result the integrand in (4.3.8) becomes

$$Q_3(\underset{\sim}{y}_{n+1}) = \exp\left\{-\frac{1}{4}(T_u + T_v)\right\} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Delta_{ij}(\underset{\sim}{y}_{n+1})\, \nu^{-(i+j)}, \tag{4.3.21}$$

where

$$\Delta_{ij}(\underset{\sim}{y}_{n+1}) = A_i(T_u)\, B_j(T_v) . \tag{4.3.22}$$

Further, using (A.1.1) through (A.1.3) of Lemma A.1 (Appendix A) we can write

$$T_u + T_v = \{(\underset{\sim}{y}_{n+1} - \underset{\sim}{\bar{\bar{y}}}_{n+1})' (\hat{Z}_u^{-1} + \hat{Z}_v^{-1}) (\underset{\sim}{y}_{n+1} - \underset{\sim}{\bar{\bar{y}}}_{n+1})\} + \{(\underset{\sim}{\hat{y}}^{(u)}_{n+1} - \underset{\sim}{\hat{y}}^{(v)}_{n+1})' (\hat{Z}_u + \hat{Z}_u)^{-1} (\underset{\sim}{\hat{y}}^{(u)}_{n+1} - \underset{\sim}{\hat{y}}^{(v)}_{n+1})\}, \tag{4.3.23}$$

where

$$\underset{\sim}{\bar{\bar{y}}}_{n+1} = (\hat{Z}_u^{-1} + \hat{Z}_v^{-1})^{-1} (\hat{Z}_u^{-1}\, \underset{\sim}{\hat{y}}^{(u)}_{n+1} + \hat{Z}_v^{-1}\, \underset{\sim}{\hat{y}}^{(v)}_{n+1}) . \tag{4.3.24}$$

Thus, finally, substituting for $(T_u + T_v)$ from (4.3.23) in (4.3.21) and going back to (4.3.8) we get

$$h_2(f_{n+1}^{(u)}, f_{n+1}^{(v)}) =$$

$$C_1[\exp\{-\tfrac{1}{4}(\hat{\underset{\sim}{y}}_{n+1}^{(u)} - \hat{\underset{\sim}{y}}_{n+1}^{(v)})'(\hat{Z}_u + \hat{Z}_v)^{-1}(\hat{\underset{\sim}{y}}_{n+1}^{(u)} - \hat{\underset{\sim}{y}}_{n+1}^{(v)})\}][|\hat{Z}_u \hat{Z}_v|^{-1/4}]$$

$$\int_{R^r}[\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\Delta_{ij}(\underset{\sim}{y}_{n+1})\,\nu^{-(i+j)}$$

$$\exp\{-\tfrac{1}{4}(\underset{\sim}{y}_{n+1} - \bar{\underset{\sim}{y}}_{n+1})'(\hat{Z}_u^{-1} + \hat{Z}_v^{-1})(\underset{\sim}{y}_{n+1} - \bar{\underset{\sim}{y}}_{n+1})\}]\; d\underset{\sim}{y}_{n+1}. \tag{4.3.25}$$

This suggests that the integral on the right hand side of (4.3.25) can be evaluated term by term and that each integral involves a bivariate polynomial in quadratic forms T_u and T_v. Let $\underset{\sim}{Y}_{n+1}$ be distributed as $N_r(\bar{\underset{\sim}{y}}_{n+1}, \{(Z_u^{-1}+Z_u^{-1})/2\}^{-1})$, then

$$h_2(f_{n+1}^{(u)}, f_{n+1}^{(v)}) =$$

$$C_1\,(2\pi)^{r/2}[|\hat{Z}_u\,\hat{Z}_v|^{-1/4}|(\hat{Z}_u^{-1} + \hat{Z}_v^{-1})/2|^{-1/2} \times$$

$$\exp\{-\tfrac{1}{4}(\hat{\underset{\sim}{y}}_{n+1}^{(u)} - \hat{\underset{\sim}{y}}_{n+1}^{(v)})'(\hat{Z}_u+\hat{Z}_v)^{-1}(\hat{\underset{\sim}{y}}_{n+1}^{(u)} - \hat{\underset{\sim}{y}}_{n+1}^{(v)})\}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\delta_{ij}\,\nu^{-(i+j)}, \tag{4.3.26}$$

where

$$\delta_{ij}\;[= E(\Delta_{ij}(\underset{\sim}{y}_{n+1}))]$$

$$= \sum_{k=1}^{i+1}\sum_{\ell=1}^{j+1}\sum_{s=1}^{k}\sum_{t=1}^{\ell}\alpha_{iks}\,\beta_{j\ell t}\,\mu_{2i-k+1,\,2j-\ell+1}\,r^{(k+\ell-s-t)}. \tag{4.3.27}$$

with $\mu_{a,b} = E(T_u^a\,T_v^b)$.

Since $\hat{Z}_u^{-1}$ and $\hat{Z}_v^{-1}$ exist, we can use the identity,

$$\hat{Z}_u(\hat{Z}_u + \hat{Z}_v)^{-1}\,\hat{Z}_v = (\hat{Z}_u^{-1} + \hat{Z}_v^{-1})^{-1}$$

and express $h_2(f_{n+1}^{(u)}, f_{n+1}^{(v)})$ of equation (4.3.26) as

$$h_2(f_{n+1}^{(u)}, f_{n+1}^{(v)}) =$$

$$C_1^* \left[\frac{4|\hat{Z}_u\,\hat{Z}_v|}{|\hat{Z}_u + \hat{Z}_v|^2}\right]^{1/4} \exp\left[-\frac{1}{4}\{(\hat{\underset{\sim}{y}}_{n+1}^{(u)} - \hat{\underset{\sim}{y}}_{n+1}^{(v)})'(\hat{Z}_u+\hat{Z}_v)^{-1}(\hat{\underset{\sim}{y}}_{n+1}^{(u)} - \hat{\underset{\sim}{y}}_{n+1}^{(v)})\}\right] \tag{4.3.28}$$

with

$$C_1^* = \left[\frac{2^{r/2}\,\Gamma(\frac{\nu+r}{2})}{\Gamma(\frac{\nu}{2})\nu^{r/2}}\right] \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \delta_{ij}\,\nu^{-(i+j)} \,. \tag{4.3.29}$$

The mixed moments $\mu_{a,b}$ involved in (4.3.27) can be obtained indirectly through the joint comulants of the quadratic forms T_u and T_v, by means of the inversion formulae of Cook (1951) (Refer D.5 of Appendix D).

So far as the evaluation of the mixed cumulants, $k_{a,b}$, ($a = 2i-k+1$, $b = 2j-\ell+1$) is concerned, we use the fact from (4.3.25) that $\underset{\sim}{Y}_{n+1}$ can be considered to be distributed according to $N_r(\bar{\bar{y}}_{n+1}, \hat{Z})$, with $\hat{Z}^{-1} = (\hat{Z}_u^{-1} + \hat{Z}_v^{-1})/2$, so that from equation (C.2.12) of Lemma C.2 (Appendix C) we have the (a,b)th cumulant of T_u and T_v given by

$$k_{a,b} = 2^{a+b-1}\,(a+b-2)!\,\{(a+b-1)\,\mathrm{tr}\,(R_u^a R_v^b) +$$

$$(a\underset{\sim}{d}_u + b\underset{\sim}{d}_v)'\,\hat{Z}\,R_u^a R_v^b\,(a\underset{\sim}{d}_u + b\underset{\sim}{d}_v) -$$

$$a\underset{\sim}{d}_u'\,\hat{Z}\,R_u^a R_v^b\,\underset{\sim}{d}_u - b\,\underset{\sim}{d}_v'\,\hat{Z}\,R_u^a R_v^b\,\underset{\sim}{d}_v\}, \qquad (4.3.30)$$

where $\underset{\sim}{d}_i = (\underset{\sim}{\bar{y}}_{n+1} - \underset{\sim}{\hat{y}}_{n+1}^{(i)})$ and $R_i = \hat{Z}^{-1}\,\hat{Z}_i$, $i = u,v$.

In connection with the conversion of $\mu_{a,b}$ to $k_{a,b}$, it may be noted that the formulae proposed to be used for this purpose are given only for the orders (a,b), such that $a+b \leq 6$. Since $(\nu^{-(i+j)})$ and the constants α_{iks} and $\beta_{j\ell t}$ decrease so rapidly with the expansion that by the time $(a+b)$ exceeds 6, the low magnitude of the factor $(\alpha_{iks}\,\beta_{j\ell t}\,\nu^{-(i+j)})$ would make the contribution from δ_{ij} negligible. Thus shows that the inversion formulae are sufficient in number for our purpose.

Finally, we point out certain important features of the function h_2 as given in equation (4.3.28). It may be noticed that, except for the factor C_1^*, h_2, like its univariate analogue, is of the same form as in the case of known Σ. Besides, even when the common covariance matrix, Σ, is replaced by its sample estimate the criterion function still takes into account the dissimilarity of the alternative probability distributions of $\underset{\sim}{Y}_{n+1}$ with respect to location and orientation which is likely to be introduced in the probability distributions of $\underset{\sim}{Y}_{n+1}$ due to the $(n+1)$th observation.

4.4 CASE 3: UNKNOWN, UNEQUAL ERROR COVARIANCE MATRICES

There is yet another situation in which it might be reasonable to assume that error considered in rival models have different covariance structures. To the list of basic assumptions $B^+(i)$ through $B^+(v)$, we therefore, append the following assumptions :

$C^+3(i)$ $\quad \Sigma_u \neq \Sigma_v, \quad u,v = 1,2,\ldots,m \ (u \neq v)$,

$C^+3(ii)$ $\quad \Sigma_u$ is not known, $u = 1,2,\ldots,m$.

Consider a random vector $\underset{\sim}{Y}_{n+1}$ on which an observation is due to be taken. From the list of assumptions if we suppress $C^+3(i)$ and $C^+3(ii)$ for the time being, then we have a situation of the type discussed in Section 4.2. Now, it has already been established in that section that if Σ_u is known then, under model u, the response vector $\underset{\sim}{Y}_{n+1}$ would be distributed according to normal distribution: $N_r(\hat{y}^{(u)}_{n+1}, W_u)$, where

$$W_u = [\Sigma_u + X^{(u)}_{n+1} \left(\sum_{k=1}^{n} \sum_{i=1}^{r} \sum_{j=1}^{r} \sigma_u^{ij} \underset{\sim}{x}^{(u)}_{ik} \underset{\sim}{x}^{(u)\prime}_{jk}\right)^{-1} X^{(u)\prime}_{n+1}] \qquad (4.4.1)$$

with σ_u^{ij} as an (i,j)th element of Σ_u^{-1} and $\underset{\sim}{x}^{(u)}_{ik}$, $X^{(u)}_{n+1}$ as given earlier in (4.2.5) and (4.2.8) , respectively.

But, for the present discussion, Σ_u has been assumed to be unknown; a situation similar to the one discussed in Section 4.3. In this case we ought to exploit the sample of observations for the estimation of the covariance matrix Σ_u, so that if $\hat{\Sigma}_u$ denotes an estimate of Σ_u, then W_u may be estimated by the matrix

$$\hat{W}_u = [\hat{\Sigma}_u + X_{n+1}^{(u)} (\sum_{k=1}^{n} \sum_{u=1}^{r} \sum_{j=1}^{r} \hat{\sigma}_u^{ij} \underset{\sim}{x}_{ik}^{(u)} \underset{\sim}{x}_{jk}^{(u)'}) X_{n+1}^{(u)'}], \quad (4.4.2)$$

where $\hat{\sigma}_u^{ij}$ denotes the (i,j)th element of $\hat{\Sigma}_u^{-1}$. Nevertheless, it may be borne in mind that the response vector, $\underset{\sim}{Y}_{n+1}$, would no longer obey a normal probability law. The theory of Section 4.2 is, therefore, not applicable in the situation thus arisen. In this case, as in Section 4.3, we use an r-variate t-distribution : $t_r(\hat{\underset{\sim}{y}}_{n+1}^{(u)}, \hat{W}_u, \nu_u)$ for $\underset{\sim}{Y}_{n+1}$, where ν_u stands for the degrees of freedom associated with the estimate $\hat{W}_u$.

Consider now two rival models, u and v (say). The p.d.f. of $\underset{\sim}{Y}_{n+1}$ under these models is given by

$$f_{n+1}^{(i)}(\underset{\sim}{y}_{n+1}/\underset{\approx}{y}) = \left[\frac{\Gamma(\frac{\nu_i+r}{2})}{\{\Gamma(\frac{1}{2})\}^r \Gamma(\frac{\nu_i}{2}) \nu_i^{r/2}}\right] |\hat{W}_i|^{-1/2} \times$$

$$[1+(\underset{\sim}{y}_{n+1} - \hat{\underset{\sim}{y}}_{n+1}^{(i)})'(\nu_i \hat{W}_i)^{-1}(\underset{\sim}{y}_{n+1} - \hat{\underset{\sim}{y}}_{n+1}^{(i)})]^{-(\nu_i+r)/2},$$

$$i = u, v. \quad (4.4.3)$$

Substituting for $f_{n+1}^{(u)}$ and $f_{n+1}^{(v)}$ from (4.4.3) in the function

$$h_3(f_{n+1}^{(u)}, f_{n+1}^{(v)}) = \int_{R^r} [f_{n+1}^{(u)}(\underset{\sim}{y}_{n+1}/\underset{\approx}{y}) \, f_{n+1}^{(v)}(\underset{\sim}{y}_{n+1}/\underset{\approx}{y})]^{1/2} \, d\underset{\sim}{y}_{n+1}$$

we obtain the affinity between $f_{n+1}^{(u)}$ and $f_{n+1}^{(v)}$ as

$$h_3(f_{n+1}^{(u)}, f_{n+1}^{(v)}) = C_2 \, |\hat{W}_u \, \hat{W}_v|^{-1/4} \int_{R^r} Q_4(\underset{\sim}{y}_{n+1}) \, d\underset{\sim}{y}_{n+1}, \quad (4.4.4)$$

where

$$C_2 = \left[\frac{\Gamma\left(\frac{\nu_u+r}{2}\right)\ \Gamma\left(\frac{\nu_v+r}{2}\right)}{\{\Gamma(\frac{1}{2})\}^{2r}\ \Gamma\left(\frac{\nu_u}{2}\right)\ \Gamma\left(\frac{\nu_v}{2}\right)(\nu_u\nu_v)^{r/2}}\right]^{1/2}, \qquad (4.4.5)$$

$$Q_4(\underset{\sim}{y}_{n+1}) = [1 + \tilde{T}_u\,\nu_u^{-1}]^{-(\nu_u+r)/4}\ [1 + \tilde{T}_v\,\nu_v^{-1}]^{-(\nu_v+r)/4}, \qquad (4.4.6)$$

and

$$\tilde{T}_i = (\underset{\sim}{y}_{n+1} - \hat{\underset{\sim}{y}}^{(i)}_{n+1})'\ \hat{W}_i^{-1}\ (\underset{\sim}{y}_{n+1} - \hat{\underset{\sim}{y}}^{(i)}_{n+1}),\ i = u,v\ . \qquad (4.4.7)$$

We notice that the integrand, $Q_4(\underset{\sim}{y}_{n+1})$, in (4.4.4) is of the same form as $Q_3(\underset{\sim}{y}_{n+1})$ in (4.3.10) (Refer Section 4.3). Therefore, in order to simplify the integrand Q_4 to a tractable form we can use the same algebra; namely, that from equations (4.3.12) through (4.3.18) of Section 4.3. This reduces the two factors of Q_4 to the forms [Refer equations (4.3.18) and (4.3.19) in Section 4.3]

$$[1 + \tilde{T}_u\,\nu_u^{-1}]^{-(\nu_u+r)/4} = \exp\left(-\tfrac{1}{4}\,\tilde{T}_u\right)\ \sum_{i=0}^{\infty} A_i(\tilde{T}_u)\ \nu_u^{-i} \qquad (4.4.8)$$

and

$$[1 + \tilde{T}_v\,\nu_v^{-1}]^{-(\nu_v+r)/4} = \exp\left(-\tfrac{1}{4}\,\tilde{T}_v\right)\ \sum_{j=0}^{\infty} B_j(\tilde{T}_v)\ \nu_v^{-j}, \qquad (4.4.9)$$

where

$$A_i(\tilde{T}_u) = \sum_{k=1}^{i+1}\ \sum_{s=1}^{k} \alpha_{iks}\ r^{k-s}\ \tilde{T}_u^{2i-k+1} \qquad (4.4.10)$$

and

$$B_j(\tilde{T}_v) = \sum_{\ell=1}^{j+1}\ \sum_{t=1}^{\ell} \beta_{j\ell t}\ r^{\ell-t}\ \tilde{T}_v^{2j-\ell+1} \qquad (4.4.11)$$

with α_{iks}'s and $\beta_{j\ell t}$'s given in Table 4.3.1.

Consequently, the integrand Q_4, assumes the form

$$Q_4(\underset{\sim}{y}_{n+1}) = \exp\{-\tfrac{1}{4}(\tilde{T}_u + \tilde{T}_v)\} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Delta_{ij}(\underset{\sim}{y}_{n+1})\, \nu_u^{-i}\, \nu_v^{-j}, \tag{4.4.12}$$

where

$$\Delta_{ij}(\underset{\sim}{y}_{n+1}) = A_i(\tilde{T}_u)\, B_j(\tilde{T}_v) .$$

Further, using (A.1.1) and (A.1.2) of Lemma A.1 (Appendix A), we have

$$\tilde{T}_u + \tilde{T}_v = \{(\underset{\sim}{y}_{n+1} - \underset{\sim}{\bar{\bar{y}}}_{n+1})'(\hat{W}_u^{-1} + \hat{W}_v^{-1})(\underset{\sim}{y}_{n+1} - \underset{\sim}{\bar{\bar{y}}}_{n+1})\} +$$

$$\{(\underset{\sim}{\hat{y}}_{n+1}^{(u)} - \underset{\sim}{\hat{y}}_{n+1}^{(v)})'(\hat{W}_u + \hat{W}_v)^{-1}(\underset{\sim}{\hat{y}}_{n+1}^{(u)} - \underset{\sim}{\hat{y}}_{n+1}^{(v)})\},$$

where

$$\underset{\sim}{\bar{\bar{y}}}_{n+1} = (\hat{W}_u^{-1} + \hat{W}_v^{-1})\ (\hat{W}_u^{-1}\, \underset{\sim}{\hat{y}}_{n+1}^{(u)} + \hat{W}_v^{-1}\, \underset{\sim}{\hat{y}}_{n+1}^{(v)}).$$

So that, finally, from (4.4.12) we obtain

$$Q_4(\underset{\sim}{y}_{n+1}) = [\exp\{-\tfrac{1}{4}(\underset{\sim}{\hat{y}}_{n+1}^{(u)} - \underset{\sim}{\hat{y}}_{n+1}^{(v)})'(\hat{W}_u + \hat{W}_v)^{-1}(\underset{\sim}{\hat{y}}_{n+1}^{(u)} - \underset{\sim}{\hat{y}}_{n+1}^{(v)})\}] \times$$

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Delta_{ij}(\underset{\sim}{y}_{n+1})\, \nu_u^{-i}\, \nu_v^{-j}$$

$$\exp\{-\tfrac{1}{4}(\underset{\sim}{y}_{n+1} - \underset{\sim}{\bar{\bar{y}}}_{n+1})'(\hat{W}_u^{-1} + \hat{W}_v^{-1})(\underset{\sim}{y}_{n+1} - \underset{\sim}{\bar{\bar{y}}}_{n+1})\} . \tag{4.4.13}$$

Having expressed the integrand in this form, it is easy now to evaluate the measure of similarity, h_3. In fact, if we now substitute for $Q_4(\underset{\sim}{y}_{n+1})$ in (4.4.4) from (4.4.13) we get

$$h_3(f_{n+1}^{(u)}, f_{n+1}^{(v)}) =$$

$$C_2[\exp\{-\tfrac{1}{4}(\hat{y}_{n+1}^{(u)} - \hat{y}_{n+1}^{(v)})'(\hat{W}_u + \hat{W}_v)^{-1}(\hat{y}_{n+1}^{(u)} - \hat{y}_{n+1}^{(v)})\}]\,|\hat{W}_u\,\hat{W}_v|^{-1/4}$$

$$\int_{R^r}\Big[\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\Delta_{ij}(y_{n+1})\,\nu_u^{-i}\,\nu_v^{-j}\times$$

$$\exp\{-\tfrac{1}{4}(y_{n+1} - \bar{\bar{y}}_{n+1})'(\hat{W}_u^{-1} + \hat{W}_v^{-1})(y_{n+1} - \bar{\bar{y}}_{n+1})\}]\,dy_{n+1}\,.$$

(4.4.14)

A look at the right hand side of (4.4.14) suggests some useful hints for the evaluation of $h_3(f_{n+1}^{(u)}, f_{n+1}^{(v)})$. Firstly, it suggests the integral involved therein can be evaluated term by term. Secondly, y_{n+1} may be considered to be normally distributed : $N_r(\bar{\bar{y}}_{n+1}, \hat{W}^{-1})$, $\hat{W}^{-1} = (\hat{W}_u^{-1} + \hat{W}_v^{-1})/2$. This makes it possible to think of each integral as an expected value of the bivariate polynomial in the quadratic forms T_u and T_v, as one observes from the expressions (4.4.10) and (4.4.11) for A_i and B_j, respectively. We thus have

$$h_3(f_{n+1}^{(u)}, f_{n+1}^{(v)}) =$$

$$C_2(2\pi)^{r/2}[|\hat{W}_u\,\hat{W}_v|^{-1/4}\,|(\hat{W}_u^{-1} + \hat{W}_v^{-1})/2|^{-1/2}]\times$$

$$\exp\{-\tfrac{1}{4}(\hat{y}_{n+1}^{(u)} - \hat{y}_{n+1}^{(v)})'(\hat{W}_u + \hat{W}_v)^{-1}(\hat{y}_{n+1}^{(u)} - \hat{y}_{n+1}^{(v)})\}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\tilde{\delta}_{ij}\nu_v^{-i}\nu_v^{-j},$$

(4.4.15)

where

$$\tilde{\delta}_{ij} = E(\tilde{\Delta}_{ij}(y_{n+1}))$$

$$= \sum_{k=1}^{i+1} \sum_{\ell=1}^{j+1} \sum_{s=1}^{k} \sum_{t=1}^{\ell} \alpha_{iks}\, \beta_{j\ell t}\, \tilde{\mu}_{2i-k+1,2j-\ell+1}\, r^{(k+\ell-s-t)}$$

and $\tilde{\mu}_{a,b} = E(\tilde{T}_u^a\, \tilde{T}_v^b)$.

Since $\hat{W}_u^{-1}$ and $\hat{W}_v^{-1}$ exist, we can use the identity,

$$\hat{W}_u(\hat{W}_u + \hat{W}_v)^{-1}\, \hat{W}_v = (\hat{W}_u^{-1} + \hat{W}_v^{-1})^{-1}$$

and, finally, write the function, h_3, as

$$h_3(f_{n+1}^{(u)}, f_{n+1}^{(v)}) =$$

$$C_2^*\left[\frac{4|\hat{W}_u\, \hat{W}_v|}{|\hat{W}_u + \hat{W}_v|^2}\right]^{1/4} \exp\left[-\tfrac{1}{4}\{(\hat{\underset{\sim}{y}}_{n+1}^{(u)} - \hat{\underset{\sim}{y}}_{n+1}^{(v)})'\,(\hat{W}_u+\hat{W}_v)^{-1}(\hat{\underset{\sim}{y}}_{n+1}^{(u)} - \hat{\underset{\sim}{y}}_{n+1}^{(v)})\}\right], \tag{4.4.16}$$

where

$$C_2^* = \left[\frac{2^r\ \Gamma(\frac{\nu_u+r}{2})\ \Gamma(\frac{\nu_v+r}{2})}{\Gamma(\frac{\nu_u}{2})\ \Gamma(\frac{\nu_v}{2})(\nu_u\nu_v)^{r/2}}\right]^{1/2} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \tilde{\delta}_{ij}\, \nu_u^{-i}\, \nu_v^{-j} \tag{4.4.17}$$

and

$$\tilde{\delta}_{ij} = \sum_{k=1}^{i+1} \sum_{\ell=1}^{j+1} \sum_{s=1}^{k} \sum_{t=1}^{\ell} \alpha_{iks}\, \beta_{j\ell t}\, \tilde{\mu}_{2i-k+1,2j-\ell+1}\, r^{(k+\ell-s-t)}. \tag{4.4.18}$$

For the evaluation of the mixed moments, $\tilde{\mu}_{a,b}$, ($a = 2i-k+1$, $b = 2j-\ell+1$) we resort to the use of inversion formulae of

Cook (1951) which express the joint moments in terms of the joint cumulants. Now, it has already been observed that $\underset{\sim}{Y}_{n+1}$ may be considered to have a normal distribution $N_r(\underset{\sim}{\bar{\bar{y}}}_{n+1}, \hat{W})$. This fact when used in Lemma C.2 of Appendix C gives the required (a,b)th cumulant $\tilde{k}_{a,b}$ as

$$\tilde{k}_{a,b} = 2^{a+b-1}\,(a+b-2)!\,\{(a+b-1)\,\mathrm{tr}\,(\tilde{R}_u^a\,\tilde{R}_v^b + (a\underset{\sim}{\bar{\bar{d}}}_u + b\underset{\sim}{\bar{\bar{d}}}_v)'\,\hat{W}\,\tilde{R}_u^a\,\tilde{R}_v^b\,(a\underset{\sim}{\bar{\bar{d}}}_u + b\underset{\sim}{\bar{\bar{d}}}_v) - a\underset{\sim}{\bar{\bar{d}}}_u'\,\hat{W}\,\tilde{R}_u^a\,\tilde{R}_v^b\,\underset{\sim}{d}_u - b\underset{\sim}{\bar{\bar{d}}}_v'\,\hat{W}\,\tilde{R}_u^a\,\tilde{R}_v^b\,\underset{\sim}{\bar{\bar{d}}}_v\}, \qquad (4.4.19)$$

where $\underset{\sim}{\bar{\bar{d}}}_i = (\underset{\sim}{\bar{\bar{y}}}_{n+1} - \underset{\sim}{\hat{y}}_{n+1}^{(i)})$, $\tilde{R}_i = \hat{W}^{-1}\,\hat{W}_i$, $i = u,v$.

It may be noted that the remark made on the adequacy of inversion formulae in Section 4.3 is applicable in the present case as well. Besides, the function h_3, given by equations (4.4.16) through (4.4.18) has the same features as its unvariate analogue in Section 3.4 of Chapter 3.

CHAPTER 5

SOME THEORY OF ESTIMATION

5.1 ESTIMATION OF MODEL PARAMETERS

A reference to the design and discrimination criteria developed in Chapters 3 and 4 brings out the important point that the physical parameters involved in the response function(s) must be estimated in order to be able to assess the level of discrimination among the rival models and to design more experiments for the purpose, if the need be. In what follows we, therefore, make a search for the appropriate objective functions which may be employed for estimation of model parameters, $\underset{\sim}{\theta}^{(u)}$, $(u = 1,2,\ldots,m)$, in various types of situations, mainly depending on our knowledge about the covariance structure of error(s).

With n data points at hand, the p_u parameters of model u (say) can be estimated effeciently if a suitable method of estimation is employed. A wide choice of such techniques is available for this purpose. However, the procedure developed here being based on the maximum likelihood estimates and the observations being normally distributed, the estimation criterion will consist of minimizing some function of the residuals. In the following we obtain specific forms of the functions suitable

for different types of situations. Of the several forms of the criterion function an appropriate one is selected on the basis of the assumptions being used in a given situation.

Consider an r-response model ($r \geq 1$), $M^{(u)}$ (say). Since the errors in each experiment have been assumed to follow a normal distribution, $N_r(\underset{\sim}{0}, \Sigma_u)$, and the errors in different experiments have been considered to be independent, the likelihood of the n sets of data, under model u, can be written as

$$L(\underset{\sim}{\theta}^{(u)}) = \{(2\pi)^r |\Sigma_u|\}^{-n/2} \exp\left[-\frac{1}{2} \operatorname{tr}\{\Sigma_u^{-1} R(\underset{\sim}{\theta}^{(u)})\}\right], \quad (5.1.1)$$

where

$$R(\underset{\sim}{\theta}^{(u)}) = \sum_{k=1}^{n} \underset{\sim}{\varepsilon}_k^{(u)} \underset{\sim}{\varepsilon}_k^{(u)'} \quad (5.1.2)$$

with

$$\underset{\sim}{\varepsilon}_k^{(u)} = \underset{\sim}{y}_k - \underset{\sim}{\eta}^{(u)}(\underset{\sim}{\xi}_k, \underset{\sim}{\theta}^{(u)}). \quad (5.1.3)$$

In the sequel, we shall use $R^{(u)}$ to denote $R(\underset{\sim}{\theta}^{(u)})$.

We now recall that in the design and discrimination criteria, we advocated the use of the m.l.e. of $\underset{\sim}{\theta}^{(u)}$. Therefore, we seek such values, $\hat{\underset{\sim}{\theta}}^{(u)}$ of $\underset{\sim}{\theta}^{(u)}$ which would maximize the likelihood function $L(\underset{\sim}{\theta}^{(u)})$. To that purpose we rather maximize the simpler function $\log_e L(\underset{\sim}{\theta}^{(u)})$, i.e., the function,

$$\ell(\underset{\sim}{\theta}^{(u)}) = -\frac{nr}{2} \log_e(2\pi) - \frac{n}{2} \log_e |\Sigma_u| - \frac{1}{2} \operatorname{tr}(\Sigma_u^{-1} R^{(u)}). \quad (5.1.4)$$

Situation 1 : If Σ_u is known, maximizing $\ell(\underset{\sim}{\theta}^{(u)})$ is equivalent to minimizing the function,

$$\phi_1(\underset{\sim}{\theta}^{(u)}) = \operatorname{tr}(\Sigma_u^{-1} R^{(u)}) \quad (5.1.5)$$

which is, therefore, the criterion for parameter estimation yielding m.l.e. of $\underset{\sim}{\theta}^{(u)}$.

Situation 2 : When the responses in each experiment are known to be independent and measured with the same precision i.e., $\Sigma_u = \sigma_u^2 I$, whether or not σ_u^2 is known the m.l.e's of the model parameters can be obtained through the minimization of

$$\phi_2(\underset{\sim}{\theta}^{(u)}) = \sum_{k=1}^{n} \underset{\sim}{\varepsilon}_k^{(u)'} \underset{\sim}{\varepsilon}_k^{(u)} \quad [= \text{tr } R(\underset{\sim}{\theta}^{(u)})]. \qquad (5.1.6)$$

Situation 3 : In yet another situation the covariance matrix Σ_u may be unknown. However, if this matrix is known to be a diagonal matrix (meaning that the responses are independent, but measured with different precisions) the criterion for parameter estimation may be obtained by stagewise maximization of $\ell(\underset{\sim}{\theta}^{(u)})$, now, given by

$$\ell(\underset{\sim}{\theta}^{(u)}) = -\frac{nr}{2}\log_e(2\pi) - \frac{n}{2}\sum_{i=1}^{r}\log_e\sigma_{u;ii} - \frac{1}{2}\sum_{i=1}^{r} R_{ii}^{(u)}/\sigma_{u;ii}, \qquad (5.1.7)$$

where $$R_{ii}^{(u)} = \sum_{k=1}^{n}\{y_{ik} - \eta_i^{(u)}(\underset{\sim}{\xi}_k, \underset{\sim}{\theta}^{(u)})\}^2. \qquad (5.1.8)$$

At first stage we seek the solution of the equation

$$\frac{\partial \ell(\underset{\sim}{\theta}^{(u)})}{\partial \sigma_{u;ii}} = -\frac{n}{2\sigma_{u;ii}} + \frac{R_{ii}^{(u)}}{2\sigma_{u;ii}^2} = 0 \ , \ i = 1,2,\ldots,r.$$

In fact, this equation has a unique solution; namely,

$$\hat{\sigma}_{u;ii} = R_{ii}^{(u)}/n. \qquad (5.1.9)$$

This expression, when used in place of $\sigma_{u;ii}$ in equation (5.1.7), reduces $\ell(\underset{\sim}{\theta}^{(u)})$ to

$$\tilde{\ell}(\underset{\sim}{\theta}^{(u)}) = -\frac{nr}{2}\log_e(2\pi) - \frac{n}{2}\sum_{i=1}^{r}\log_e(R_{ii}^{(u)}/n) - \frac{1}{2}\sum_{i=1}^{r}\{R_{ii}^{(u)}/(R_{ii}^{(u)}/n)\}$$

$$= \frac{nr}{2}\{\log_e(\frac{n}{2\pi}) - 1\} - \frac{n}{2}\sum_{i=1}^{r}\log_e R_{ii}^{(u)}. \quad (5.1.10)$$

Thus the function $\tilde{\ell}(\underset{\sim}{\theta}^{(u)})$ can be maximized with respect to $\underset{\sim}{\theta}^{(u)}$ by minimizing

$$\phi_3(\underset{\sim}{\theta}^{(u)}) = \frac{n}{2}\sum_{i=1}^{r}\log_e R_{ii}^{(u)}. \quad (5.1.11)$$

The case of correlated responses, which are measured with different accuracies, can be dealt with similarly. We, as a first step, maximize $\ell(\underset{\sim}{\theta}^{(u)})$ of equation (5.1.4) with respect to the covariance matrix Σ_u. This amounts to solving the matrix equation

$$\frac{\partial \ell}{\partial \Sigma_u} = -\frac{n}{2}\Sigma_u^{-1} + \frac{1}{2}\Sigma_u^{-1}R^{(u)}\Sigma_u^{-1} = 0_{r\times r}, \quad (5.1.12)$$

where $r \times r$ matrix 0 denotes a matrix with all zero entries. This equation can be rewritten as

$$\Sigma_u^{-1} = \frac{1}{n}\Sigma_u^{-1}R^{(u)}\Sigma_u^{-1},$$

so that on premultiplying and postmultiplying both sides by Σ_u, we get

$$\Sigma_u = \frac{1}{n}R^{(u)} = \hat{\Sigma}_u \text{ (say)}. \quad (5.1.13)$$

As a result, $\ell(\underset{\sim}{\theta}^{(u)})$ assumes the form

$$\tilde{\ell}(\underset{\sim}{\theta}^{(u)}) = -\frac{nr}{2}\log_e(2\pi) - \frac{n}{2}\log_e|R^{(u)}/n| - \frac{n}{2}\operatorname{tr}(R^{(u)^{-1}}R^{(u)})$$

$$= -\frac{nr}{2}\{\log_e(2\pi)+1\} + \frac{n}{2}\log_e n - \frac{n}{2}\log_e|R^{(u)}|. \tag{5.1.14}$$

Therefore, when Σ_u is a general unknown covariance matrix, the m.l.e. of $\underset{\sim}{\theta}^{(u)}$ can be obtained by minimizing

$$\phi_4(\underset{\sim}{\theta}^{(u)}) = |R(\underset{\sim}{\theta}^{(u)})|. \tag{5.1.15}$$

5.2 ESTIMATION OF VARIANCE(S) COVARIANCES OF ERROR(S)

So far as the design of experiments for model discrimination is concerned, it may be noted that the estimation of certain variances or covariance matrices is equally important. We, first, obtain an estimate of the covariance matrix, Σ_u. In fact, one type of estimator of Σ_u is already known to us from Section 5.1; namely, the solution to the equation (5.1.12). So that once the m.l.e. of $\underset{\sim}{\theta}^{(u)}$ have been determined the m.l.e. of Σ_u assumes the form

$$\hat{\Sigma}_u = \frac{1}{n}\sum_{k=1}^{n} \underset{\sim}{e}_k^{(u)}\, \underset{\sim}{e}_k^{(u)'}, \tag{5.2.1}$$

where

$$\underset{\sim}{e}_k^{(u)} = \{\underset{\sim}{y}_k - \underset{\sim}{\eta}^{(u)}(\underset{\sim}{\xi}_k, \hat{\underset{\sim}{\theta}}^{(u)})\}. \tag{5.2.2}$$

We seek an unbiased estimate of Σ_u. Therefore, in what follows we examine Σ for unbiasedness and remove the bias, if any.

From (5.2.1), we have

$$E^{(u)}(\hat{\Sigma}_u) = \frac{1}{n}\sum_{k=1}^{n} E^{(u)}\,(\underset{\sim}{e}_k^{(u)}\ \underset{\sim}{e}_k^{(u)\prime}). \qquad (5.2.3)$$

Now, assuming that $\hat{\underset{\sim}{\theta}}^{(u)}$ does not differ much from the true value $\underset{\sim}{\theta}^{(u)}$, we can write, by Taylor's expansion of the response function $\underset{\sim}{\eta}^{(u)}(\underset{\sim}{\xi}_k, \underset{\sim}{\theta}^{(u)})$ around $\hat{\underset{\sim}{\theta}}^{(u)}$,

$$\underset{\sim}{y}_k = \underset{\sim}{\eta}^{(u)}(\underset{\sim}{\xi}_k, \hat{\underset{\sim}{\theta}}^{(u)}) + X_k^{(u)}(\underset{\sim}{\theta}^{(u)} - \hat{\underset{\sim}{\theta}}^{(u)}) + \underset{\sim}{\varepsilon}_k^{(u)},$$

so that

$$\underset{\sim}{e}_k^{(u)} = X_k^{(u)}(\underset{\sim}{\theta}^{(u)} - \hat{\underset{\sim}{\theta}}^{(u)}) + \underset{\sim}{\varepsilon}_k^{(u)}\ . \qquad (5.2.4)$$

Besides, it has already been seen in Section 5.1 that when Σ_u is known, the m.l.e. $\hat{\underset{\sim}{\theta}}^{(u)}$ of $\underset{\sim}{\theta}^{(u)}$ is supposed to minimize the objective function of the form

$$\phi_4(\underset{\sim}{\theta}^{(u)}) = \Psi(R^{(u)}).$$

This implies that at this minimum the first derivative of ϕ_4 must vanish. That is to say,

$$\Big[\sum_{i=1}^{r}\sum_{j=1}^{r}\Big(\frac{\partial \Psi}{\partial R_{ij}^{(u)}}\frac{\partial R_{ij}^{(u)}}{\partial \theta_s^{(u)}}\Big)\Big]_{\underset{\sim}{\theta}^{(u)}=\hat{\underset{\sim}{\theta}}^{(u)}} = 0. \qquad (5.2.5)$$

We now recall that

$$R_{ij}^{(u)} = \sum_{k=1}^{n} \varepsilon_{ik}\,\varepsilon_{jk}\ , \qquad (5.2.6)$$

where $\varepsilon_{ik} = y_{ik} - \eta_i^{(u)}(\underset{\sim}{\xi}_k, \underset{\sim}{\theta}^{(u)})$.

On differentiating (5.2.6) with respect to $\theta_s^{(u)}$, we get

$$\frac{\partial R_{ij}^{(u)}}{\partial \theta_s^{(u)}} = - \sum_{k=1}^{n} (\varepsilon_{ik}\, \tau_{jks} + \varepsilon_{jk}\, \tau_{iks}), \qquad (5.2.7)$$

where

$$\tau_{iks} = \frac{\partial \eta_i^{(u)}\, (\underset{\sim}{\xi}_k, \underset{\sim}{\theta}^{(u)})}{\partial \theta_s^{(u)}} . \qquad (5.2.8)$$

Using (5.2.8) and the symmetry of $R^{(u)}$, equation (5.2.5) results into the relation

$$\sum_{k=1}^{n} X_k^{(u)'}\, Z^{(u)}\, \underset{\sim}{e}_k^{(u)} = \underset{\sim}{0} \qquad (5.2.9)$$

where $X_k^{(u)}$ and $Z^{(u)}$ are given by

$$X_k^{(u)} = \left[\frac{\partial \underset{\sim}{\eta}^{(u)}\, (\underset{\sim}{\xi}_k, \underset{\sim}{\theta}^{(u)})}{\partial \underset{\sim}{\theta}^{(u)}}\right]_{\underset{\sim}{\theta}^{(u)} = \hat{\underset{\sim}{\theta}}^{(u)}} \qquad (5.2.10)$$

and

$$Z^{(u)} = \left[\frac{\partial \Psi(R^{(u)})}{\partial R^{(u)}}\right]_{\underset{\sim}{\theta}^{(u)} = \hat{\underset{\sim}{\theta}}^{(u)}} \qquad (5.2.11)$$

It may be noted from (5.1.5) that when Σ_u is known,

$$\Psi(R^{(u)}) = \frac{1}{2} \operatorname{tr} (\Sigma_u^{-1}\, R^{(u)}), \qquad (5.2.12)$$

so that

$$Z^{(u)} = \frac{1}{2}\, \Sigma_u^{-1} , \qquad (5.2.13)$$

while in case of Σ_u as an unknown general covariance matrix, we have from (5.1.14)

$$\Psi(R^{(u)}) = \frac{n}{2} \log_e |R^{(u)}|. \qquad (5.2.14)$$

Accordingly, $Z^{(u)}$ in this case will assume the form

$$Z^{(u)} = \frac{n}{2}\, (R^{(u)})^{-1}. \qquad (5.2.15)$$

Substituting for $\underset{\sim}{e}_k^{(u)}$ from (5.2.4) in (5.2.9) we get

$$\sum_{k=1}^{n} X_k^{(u)'} Z^{(u)} [\underset{\sim}{\varepsilon}_k - X_k^{(u)}(\underset{\sim}{\theta}^{(u)} - \underset{\sim}{\hat{\theta}}^{(u)})] = \underset{\sim}{0}$$

which in turn gives

$$\underset{\sim}{\theta}^{(u)} - \underset{\sim}{\hat{\theta}}^{(u)} = H^{-1} \sum_{\ell=1}^{n} X_\ell^{(u)'} Z^{(u)} \underset{\sim}{\varepsilon}_\ell \quad , \tag{5.2.16}$$

where

$$H = \sum_{k=1}^{n} X_k^{(u)'} Z^{(u)} X_k^{(u)} \quad . \tag{5.2.17}$$

Going back to (5.2.4) and using the resulting expression for $(\underset{\sim}{\theta}^{(u)} - \underset{\sim}{\hat{\theta}}^{(u)})$ from (5.2.16) we get

$$\underset{\sim}{e}_k^{(u)} = \underset{\sim}{\varepsilon}_k - X_k^{(u)} H^{-1} \sum_{\ell=1}^{n} X_\ell^{(u)'} Z^{(u)} \underset{\sim}{\varepsilon}_\ell \quad . \tag{5.2.18}$$

This expression for $\underset{\sim}{e}_k^{(u)}$ is now utilized for evaluation of $E^{(u)}(\underset{\sim}{e}_k^{(u)} \underset{\sim}{e}_k^{(u)'})$. In fact,

$$E^{(u)}(\underset{\sim}{e}_k^{(u)} \underset{\sim}{e}_k^{(u)'}) = \Sigma_u - X_k^{(u)} H^{-1} X_k^{(u)'} Z^{(u)} \Sigma_u - \Sigma_u Z^{(u)} X_k^{(u)} H^{-1} X_k^{(u)'} + X_k^{(u)} H^{-1} (\sum_{\ell=1}^{n} X_\ell^{(u)'} Z^{(u)} \Sigma_u Z^{(u)'} X_\ell^{(u)}) H^{-1} X_k^{(u)'} . \tag{5.2.19}$$

Consequently, from (5.2.3) we have

$$E(\hat{\Sigma}_u) = \Sigma_u - \frac{1}{n} (\sum_{k=1}^{n} X_k^{(u)} H^{-1} X_k^{(u)'}) Z^{(u)} \Sigma_u - \frac{1}{n} \Sigma_u Z^{(u)} (\sum_{k=1}^{n} X_k^{(u)} H^{-1} X_k^{(u)'}) + \frac{1}{n} \sum_{k=1}^{n} X_k^{(u)} H^{-1} (\sum_{\ell=1}^{n} X_\ell^{(u)'} Z^{(u)} \Sigma_u Z^{(u)'} X_\ell^{(u)}) H^{-1} X_k^{(u)'} . \tag{5.2.20}$$

It may be noted at this stage that the relation (5.2.20) between $\hat{\Sigma}_u$ and Σ_u remains unchanged if $Z^{(u)}$ is multiplied by a constant, since H would be multiplied by the same constant. Let us suppose for the time being that Σ_u is known. Using the fact stated above, if we substitute $Z^{(u)} = \Sigma_u^{-1}$ in (5.2.20), we obtain

$$E(\hat{\Sigma}_u) = \Sigma_u - \frac{2}{n}\sum_{k=1}^{n} X_k^{(u)} H^{-1} X_k^{(u)'} + \frac{1}{n}\sum_{k=1}^{n} X_k^{(u)} H^{-1} H H^{-1} X_k^{(u)'}$$

$$= \Sigma_u - \frac{1}{n}\sum_{k=1}^{n} X_k^{(u)} H^{-1} X_k^{(u)'} . \qquad (5.2.21)$$

This shows that $E^{(u)}(\hat{\Sigma}_u)$ is smaller than $\hat{\Sigma}_u$ and hence is a biased estimator.

We now proceed to remove the bias in Σ_u, in an r-response case. In fact, we seek an estimate of Σ_u which is some multiple of Σ_u so that

$$E^{(u)}(\hat{\Sigma}_u) = \nu_u \Sigma_u . \qquad (5.2.22)$$

To that purpose we utilize the equation (5.2.21). This equation, no doubt, is based on the assumption that Σ_u is known but its use in case of unknown Σ_u can be justified on the ground that in the latter case $Z^{(u)}$ is proportional to $(R^{(u)})^{-1}$ and the m.l.e. of Σ_u is proportional to $R^{(u)}$ so that $Z^{(u)}$ can be taken proportional to Σ_u^{-1}. Besides, the multiplication of $Z^{(u)}$ by a constant does not make any difference so far as (5.2.21) is concerned, since H would be multiplied by the same constant. The equation (5.2.21), therefore, remains valid even in the case when Σ_u is not known. Substituting for $E^{(u)}(\hat{\Sigma}_u)$ from

(5.2.21) in (5.2.22) we obtain

$$\nu_u \Sigma_u = \Sigma_u - \frac{1}{n} \sum_{k=1}^{n} X_k^{(u)} H^{-1} X_k^{(u)'} .$$

Post multiplying both sides by Σ_u^{-1} we get

$$\nu_u I_r = I_r - \frac{1}{n} \sum_{k=1}^{n} X_k^{(u)} H^{-1} X_k^{(u)'} \Sigma_u^{-1} .$$

Taking trace on both sides and using the fact that $(X_k^{(u)} H^{-1} X_k^{(u)'} \Sigma_u^{-1})$ is a square matrix, we, finally, have

$$\nu_u r = r - \frac{1}{n} tr \left[H^{-1} \left(\sum_{k=1}^{n} X_k^{(u)'} \Sigma_u^{-1} X_k^{(u)}\right)\right]$$

$$= r - \frac{1}{n} tr(H^{-1}H)$$

$$= r - \frac{p_u}{n} .$$

So that,

$$\nu_u = \left[1 - \left(\frac{p_u}{nr}\right)\right]. \qquad (5.2.23)$$

Consequently, an unbiased estimate of Σ is given by

$$\hat{\Sigma}_u = \left(n - \frac{p_u}{r}\right)^{-1} \sum_{k=1}^{n} \underset{\sim}{e}_k^{(u)} \underset{\sim}{e}_k^{(u)'} . \qquad (5.2.24)$$

We have thus seen that the m.l.e. of Σ_u is biased by the factor $\left(1 - \frac{p_u}{nr}\right)^{-1}$; that an unbiased estimator of the covariance matrix, Σ_u, can be obtained through the moment matrix of residuals; and that the number of degrees of freedom associated with this estimation is the number of observations per response less the average number of parameters per equation.

Corollary

In the univariate case, we know from (5.1.13) that the m.l.e. of σ_u^2 is given by

$$\hat{\sigma}_u^2 = \frac{1}{n} \sum_{k=1}^{n} (\varepsilon_k^{(u)})^2.$$

But, once the m.l.e.'s of $\underset{\sim}{\theta}^{(u)}$ have been secured, this estimate assumes the form

$$\hat{\sigma}_u^2 = \frac{1}{n} \sum_{k=1}^{n} (e_k^{(u)})^2, \qquad (5.2.25)$$

where

$$e_k^{(u)} = [y_k - \eta^{(u)} (\underset{\sim}{\xi}_k, \hat{\underset{\sim}{\theta}}^{(u)})] .$$

Further, using the algebra similar to that of the multivariate case, it can be shown that this estimate is biased and that the bias can be removed by multiplying $\hat{\sigma}_u^2$ of (5.2.25) by $(n-p_u)^{-1}$. Thus an unbiased estimate of σ_u^2 is given by

$$\hat{\sigma}_u^2 = \frac{\sum_{k=1}^{n} (e_k^{(u)})^2}{n-p_u} \qquad (5.2.26)$$

with ν_u [$= (n-p_u)$] degrees of freedom.

5.3 ESTIMATION OF Σ_u AND σ_u^2 UNDER DIFFERENT TYPES OF ASSUMPTIONS

Multivariate Models : From the discussions in Chapter 4 we find that so far as the covariance matrix of errors, Σ_u, under model u is concerned, it is in two cases that we need its estimate.

Case (i) : When $\Sigma_u = \Sigma$, $u = 1,2,\ldots,m$, but Σ is unknown. In this case the information about the equality of covariance matrices must be incorporated. To that purpose we pool the estimates $\hat{\Sigma}_u$'s, given in (5.2.24), and obtain an estimate of the common covariance matrix Σ as

$$\hat{\Sigma} = \frac{\sum_{u=1}^{m} \sum_{k=1}^{n} (\underset{\sim}{e}_k^{(u)} \underset{\sim}{e}_k^{(u)'})}{\sum_{u=1}^{m} (n - \frac{p_u}{r})} \qquad (5.3.1)$$

with $\nu \left[= \sum_{u=1}^{m} (n - \frac{p_u}{r})\right]$ degrees of freedom. Since in this case Σ_u's are assumed to be equal, the estimate given in (5.3.1) provides us with an estimate for each Σ_u.

Case (ii) : In another case, we assumed in Chapter 4 that $\Sigma_u \neq \Sigma_v$, $u,v = 1,2,\ldots,m$ $(u \neq v)$ and that none of these covariance matrices is known. The estimate of Σ_u will then be simply given by

$$\hat{\Sigma}_u = \frac{\sum_{k=1}^{n} \underset{\sim}{e}_k^{(u)} \underset{\sim}{e}_k^{(u)'}}{(n - \frac{p_u}{r})} \qquad (5.3.2)$$

and has $\nu_u \left[= (n - \frac{p_u}{r})\right]$ degrees of freedom.

Univariate Models : While designing an experiment for a single response system we have seen in Chapter 3 that the estimate of the variance of errors under model u is needed in two cases

Case (i) : When $\sigma_u^2 = \sigma^2$, $u = 1,2,\ldots,m$, but σ^2 is unknown. Because of the homogeneity of variances, σ_u^2's, an estimate of the

common variance, σ^2, can be obtained by pooling the estimates given in (5.2.26). In fact such an estimate can be specified as

$$\hat{\sigma}^2 = \frac{\sum_{u=1}^{m} \sum_{k=1}^{n} (e_k^{(u)})^2}{(nm - \sum_{u=1}^{n} p_u)} \qquad (5.3.4)$$

with $\nu\,[= (nm - \sum_{u=1}^{n} p_u)]$ degrees of freedom. In the present case it is this estimate which is to be used as an estimate for each σ_u^2.

Case (ii) : $\sigma_u^2 \neq \sigma_v^2$, $u,v = 1,2,\ldots,m$ $(u \neq v)$ with all σ_u^2's as unknown quantities. The heterogeneity of variances in this case does not allow the pooling of the estimates $\hat{\sigma}_u^2$'s. Therefore, no further modification is required and the estimate given in (5.2.26) is used in the design criterion pertinent to this case.

CHAPTER 6

APPLICATION EXAMPLES

6.1 THE SCHEME FOR IMPLEMENTATION OF THE DISCRIMINATION PROCEDURE

The algorithms developed in the preceding sections for discriminating among rival models are now implemented in different types of situations. The technique proposed here is not only demonstrated through its application to linear and nonlinear models in r-response ($r \geq 1$) situations, but is also compared with some of the procedures proposed elsewhere. As far as the data are concerned we resort to Monte Carlo method. The model that we use for simulating observations would mostly be one of the rivals. Once the data have been obtained this role of the model, assumed to be true, is suspended for the time being. The said model would then be considered as one of the competing models and would, therefore, be treated on the same footing as the other rival models. The next step in the proposed procedure would be the estimation of the parameters of all the models under consideration. This would enable us to use the discrimination criterion so as to assess the current position; it would be natural to expect the same model to show up as the best model which was used for simulating the data.

The failure to discriminate at this stage would mean the necessity of designing an additional point, with more discriminatory power, through an appropriate design criterion. Another value of the response can now be generated and used in the assessment of discrimination at this stage. This would be continued till the given procedure allows us to stop by declaring a model from amongst the rivals as the most adequate. The entire scheme for implementation of the proposed discrimination procedure is shown in the flow diagram of Figure 6.1.1. All computations have been done on the Computer System : DEC 1090, using the programming language : FORTRAN .

6.1.1 Simulation of Data

As a first step we seek the data which can be used initially for discriminating among the given set of competing models. While dealing with a real life system the data can be obtained by carrying out a series of experiments. But, in the absence of such a system the experiments can be simulated on a computer. While simulation of experiments is indispensible in such a situation it may be advisible to employ it in others because of the advantages associated with its use. For example, even if it is possible to actually conduct the experiment it may be wise to, first, test a method on computer simulated experiments and then apply it to the real life problem. This way we can determine economically whether the method is likely to be successful in actual practice Besides, in order to test a method for immunity to sampling

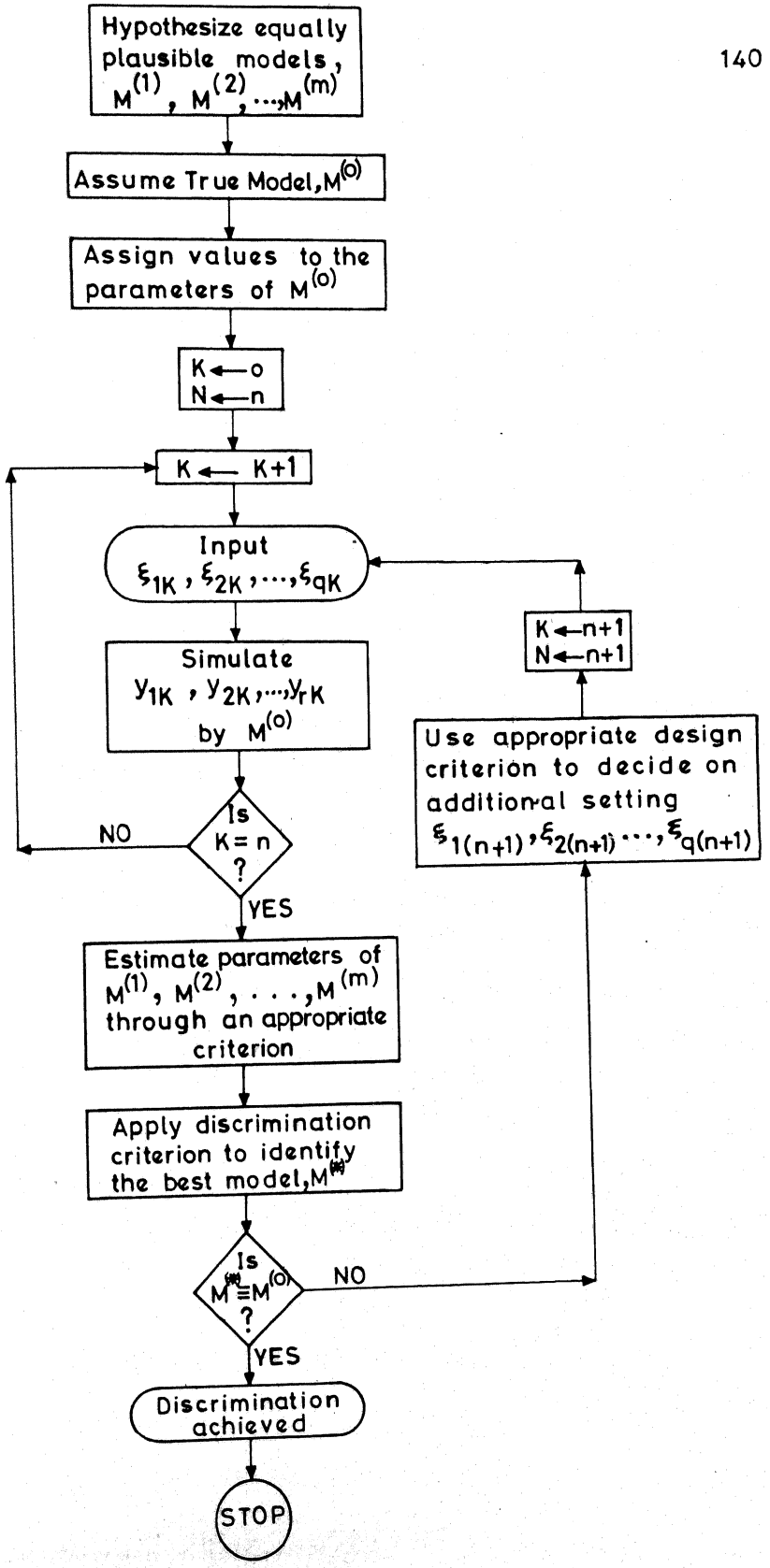

FIGURE.6.1.1 Scheme for illustration of sequential discrimination.

fluctuations it is essential to repeat the whole series of experiments many times, which may be an expensive job if one plans to work on a real life system. This object, too, can be achieved through simulation on a computer at a far less cost. With all these considerations we adopt simulation rather than experimentation for acquiring data, initially, as well as at later stages if a sequential scheme of discrimination is followed.

Now, an experiment, from our point of view, is merely a device which yields value(s) of the response(s) when an input of the controllable variable(s) is fed into it. We can, therefore, simulate it through an algorithm which computes the value(s) of r response(s) $\underset{\sim}{Y}$ for the given sets of values of $\underset{\sim}{\xi}$, using the formulae,

$$\underset{\sim}{Y} = \underset{\sim}{\eta}^{(0)} (\underset{\sim}{\xi}, \underset{\sim}{\theta}^{(0)}) + \underset{\sim}{\varepsilon} , \qquad (6.1.1)$$

where $\underset{\sim}{\eta}^{(0)}$ denotes the functional form of the model, $M^{(0)}$ (say), assumed to be the true model for the hypothetical system, $\underset{\sim}{\theta}^{(0)}$ represents the set of values assigned to the parameters of $M^{(0)}$, and the error term, $\underset{\sim}{\varepsilon}$, consists of pseudorandom number(s) drawn from $N_1(0,\sigma^2)$ if the simulation corresponds to a single response system, and from $N_r(\underset{\sim}{0},\Sigma)$, if the system under consideration is multiresponse.

6.1.2 Estimation of Model Parameters

Whether the data are acquired from an experiment or through simulation, as we have planned to do in the present study, the

parameters of all the competing models must be estimated before we proceed to discriminate. If the discrimination is to be carried out sequentially this task is required to be performed repeatedly in design and analysis. The estimates of the physical parameters as well as those of the other quantities obtained at a previous stage must be updated after each additional set of value(s) of the response(s) is appended to the data at hand. This makes estimation as one of the important steps in the process of model discrimination. If the aim is to estimate the model parameters, $\underset{\sim}{\theta}^{(u)}$, then an estimation criterion from amongst the proposed ones; namely, ϕ_1, ϕ_2, ϕ_3, and ϕ_4, given by (5.1.5), (5.1.6), (5.1.11), and (5.1.15), respectively, may be exploited; the choice depending on the assumption about the error covariance structure being met in a given situation. When the estimates of other quantities such as, Σ_u , Σ , σ_u^2 , and σ^2 are required, the formulae in (5.3.2) , (5.3.1), (5.2.26), and (5.3.4), respectively, provide us with efficient estimators.

6.1.3 Discrimination among Models and Design of Additional Experiments

Having obtained the required estimates, the selection of the best model may be attempted. This can be done by means of discrimination index (2.4.1) which utilizes the distances, $K_n^{(u)}$, $u = 1,2,\ldots,m$, from equation (2.5.25) or (2.5.27), according as the system is uniresponse or multiresponse. If the discrimination does not seem to have been achieved at a

certain stage, more experiments may be designed by maximizing the criterion function $\phi(\underset{\sim}{\xi}_{n+1})$ of (2.4.5), formed through a suitable function from amongst h_1, h_2, and h_3, given in (3.2.22), (3.3.32), and (3.4.20), respectively, if the system is uniresponse and from amongst those given in (4.2.30), (4.3.28), and (4.4.16), if the system is multiresponse.

6.1.4 Optimization of the Criterion Functions

While implementing the discrimination procedure laid down in this study the investigator will be encountered with the problem of optimizing a function, at two stages. Firstly, when the estimates of the model parameters are required to be obtained through an estimation criterion function and secondly, when a new experiment is to be set up through a design criterion. Posed in the form of ϕ_1, ϕ_2, ϕ_3, and ϕ_4 the parameter estimation problem appears simply the minimization problem in the parameter space R_θ, in which the y's and ξ's are the given numbers and θ's are the variables. On the other hand the design problem, proposed to be solved through some appropriate criterion function, is a maximization problem in the space, R_ξ, i.e., the operability range(s) of the input variable(s), where y's and θ's are treated as fixed numbers and $\underset{\sim}{\xi}_{n+1}$ as the variables. Many of the tools of deterministic optimization can be successfully brought to bear on the problem of optimization. Because of the difficulty involved in obtaining analytical derivatives or the time required for computing numerical derivatives of the objective function in

action especially, when the models are nonlinear and involve more than one function we prefer using a derivative free method.

Now, while estimating the parameters in a mechanistic model the permissible space of the physical parameters may be known. This asks for constrained minimization of the function involved, with θ's confined to the parameter space. To that purpose in this work we employ a multivariable constrained optimization routine, "Golden Complex Search" [Source : R.R. Hughes, Univ. of Wisconsin, Madison; Language : FORTRAN], based on the 'complex' method of Box (1965). In addition to the basic algorithm of Box, this routine incorporates the following provisions (i) golden search between the reflected point and the discarded point, (ii) random generation of new points, if the collapse of the complex threatens, (iii) the random restart(s) with a smaller complex centred on the trial optimum (optima), (iv) the weighting centroid calculations by factors proportional to some power of the differences between the objective function values for the points and that for the worst point in the simplex.

As regards the selection of optimal experimental conditions, the choice is generally not unrestricted. Therefore, searching for the maximum of ϕ involves constrained optimization, with the input variables confined to a bounded feasible region. In such a case, the grid search technique has been used, as the number of variables in the applications discussed here is small, being, 1 or 2. It may, however, be pointed out that when there

are 3 or more variables, the number of points on the grid may become too large. In such cases the maximum may be realized by using the above mentioned constrained optimization routine.

6.2 DISCRIMINATION AMONG UNIVARIATE MODELS

6.2.1 Example : Linear models; Known, Equal Error Variances

We start with a simple example of discriminating among univariate linear models. Consider a situation in which the data are simulated through the polynomial, $(1 + \xi + \xi^2)$, and the most adequate model is to be chosen from amongst the four models; namely,

$$M^{(1)} : \eta^{(1)} = \theta_1^{(1)} \xi \quad ,$$

$$M^{(2)} : \eta^{(2)} = \theta_1^{(2)} + \theta_2^{(2)} \xi \quad ,$$

$$M^{(3)} : \eta^{(3)} = \theta_1^{(3)} + \theta_2^{(3)} \xi + \theta_3^{(3)} \xi^2 \quad ,$$

$$M^{(4)} : \eta^{(4)} = \theta_1^{(4)} \xi + \theta_2^{(4)} \xi^2 \quad .$$

Initially, five settings of the independent variable, ξ, are chosen and a sample of five values of the response, Y, is formed through the relation

$$y = 1 + \xi + \xi^2 + \varepsilon, \qquad (6.2.1)$$

where ε has been assumed to be distributed as $N_1(0,1)$. The results obtained are presented through runs 1 to 5 in Table 6.2.2. Based on these data, the estimates of the parameters $\underset{\sim}{\theta}^{(1)}$, $\underset{\sim}{\theta}^{(2)}$, $\underset{\sim}{\theta}^{(3)}$, and $\underset{\sim}{\theta}^{(4)}$ are secured through the criteria function.

$\phi_1(\underset{\sim}{\theta}^{(u)})$, and are presented in Table 6.2.1. This table also shows the estimates of other quantities, involved in the distance, $K_n^{(u)}$, given by (2.5.25). These estimates are calculated from the samples, supposed to have been drawn from the populations corresponding to the rival models. Having assumed that at the initial stage $D_4^{(u)} = 0.25$, u = 1,2,3,4,the use of these estimates further enables us to determine the values of $D_5^{(u)}$, u = 1,2,3,4 as shown in Table 6.2.2. A comparison of the values, 0.6981, 0.1687, 0.0197, 0.1135, of the discrimination index, based on the preliminary data of 5 points not only brings to light the fact that $M^{(3)}$ is the best model (with the least value, 0.0204, of the discrimination index)but also indicates that $M^{(1)}$, with $D_5^{(1)} = 0.6981$ is the worst model for the data simulated through a general quadratic. Thus the proposed discrimination criterion picks up the correct model, i.e., the one which generated the data.

In order to see the trend of discrimination index and the efficiency of the discriminatory design criterion, max $\phi(\xi_{n+1})$, in sequential discrimination we design some additional points using (3.2.23). To that purpose $\underset{\sim}{\xi}$ is constrained to lie in the interval, $\{0 \leq \xi \leq 4\}$, and a search is always made over the grid, {0.0(0.2) 4.0}, of 21 points. The results are obtained in Table 6.2.2. It can be seen through this table that just another point, $\xi_6 = 4.0$, added to the initial data pulls the value of $D_6^{(3)}$ down to 0.0005. On the other hand,the value of $D_6^{(1)}$ rises to 0.9332. At this stage itself it seems as if there is no

Table 6.2.1
Estimation of model parameters at different stages

Stage	Model u	Estimates of parameters $\theta_1^{(u)}$	$\theta_2^{(u)}$	$\theta_3^{(u)}$
1	1	4.5056		
	2	-1.4005	4.9717	
	3	0.6596	0.8492	1.0310
	4	1.4254	0.9240	
2	1	4.7374		
	2	-1.6768	5.2481	
	3	0.6672	0.8222	1.0413
	4	1.3927	0.9382	
3	1	4.6652		
	2	-2.0063	5.3097	
	3	0.6748	1.0544	0.9805
	4	1.6386	0.8741	
4	1	4.5604		
	2	-2.5081	5.4031	
	3	0.6701	0.8691	1.0288
	4	1.4530	0.9225	
5	1	4.7620		
	2	-2.7862	5.6423	
	3	0.7027	0.7601	1.0740
	4	1.3660	0.9651	

Table 6.2.2

Sequential discrimination among polynomial models:
the present and the Box-Hill procedures
(True model : $M^{(3)}$)

Run k	Input Variable ξ_k	Response y_k	Discrimination Criteria $D_k^{(1)}$ ($P_k^{(1)}$)	$D_k^{(2)}$ ($P_k^{(2)}$)	$D_k^{(3)}$ ($P_k^{(3)}$)	$D_k^{(4)}$ ($P_k^{(4)}$)
1	0.0	1.431				
2	1.0	2.575				
3	2.0	7.540				
4	3.0	11.765				
5	4.0	20.442	0.6981	0.1687	0.0197	0.1135
	(4.0)	(20.442)	(0.00)	(0.00)	(0.66)	(0.33)
6	4.0	20.692	0.9332	0.0443	0.0005	0.0220
	(0.0)	(1.837)	(0.00)	(0.00)	(0.88)	(0.12)
7	1.6	5.732	0.9899	0.0077	0.0000	0.0024
	(0.0)	(0.140)	(0.00)	(0.00)	(0.75)	(0.25)
8	1.6	4.124	0.9987	0.0011	0.0000	0.0002
	(0.0)	1.686	(0.00)	(0.00)	(0.90)	(0.10)
9	4.0	21.623	0.9998	0.0002	0.0000	0.0001
	(0.0)	(1.714)	(0.00)	(0.00)	(0.97)	(0.03)

necessity of designing more points. However, 4 additional values of ξ are acquired through the design-criterion function (3.2.23) so as to check the validity of the claim made at the previous (sixth) run and see the trend of the criterion. This is depicted through Figure 6.2.1. The graphs in this figure show that the value of $D_k^{(3)}$ keeps dropping fast as more and more discriminatory points are added to the data. It also points out that $M^{(4)}$ is the closest rival of $M^{(3)}$. Besides, through a close look at the values of discrimination index in Table 6.2.2, one may notice that the models have been ordered according to their appropriateness. Such an information may prove to be useful when cost of using a model is a consideration and it may be cheaper to use a model other than the one selected through the criterion. In this case one would naturally look for the second best.

Box and Hill (1967) have considered the same problem of discriminating among four polynomial models. The results obtained by them are presented through the bracketed values in Table 6.2.2. Looking at these values one notices that after 5 preliminary data points the posterior probability no doubt indicates that $M^{(3)}$ with 0.66 as its probability, is the best model but not as clearly as the discrimination index does. As one moves further one notices that the altitude which $P_{5+}^{(3)}$ attains at the sixth run; being 0.88, suddenly drops to 0.75. This leaves one double minded as to whether to declare this model as the best, for it might drop further at the next run.

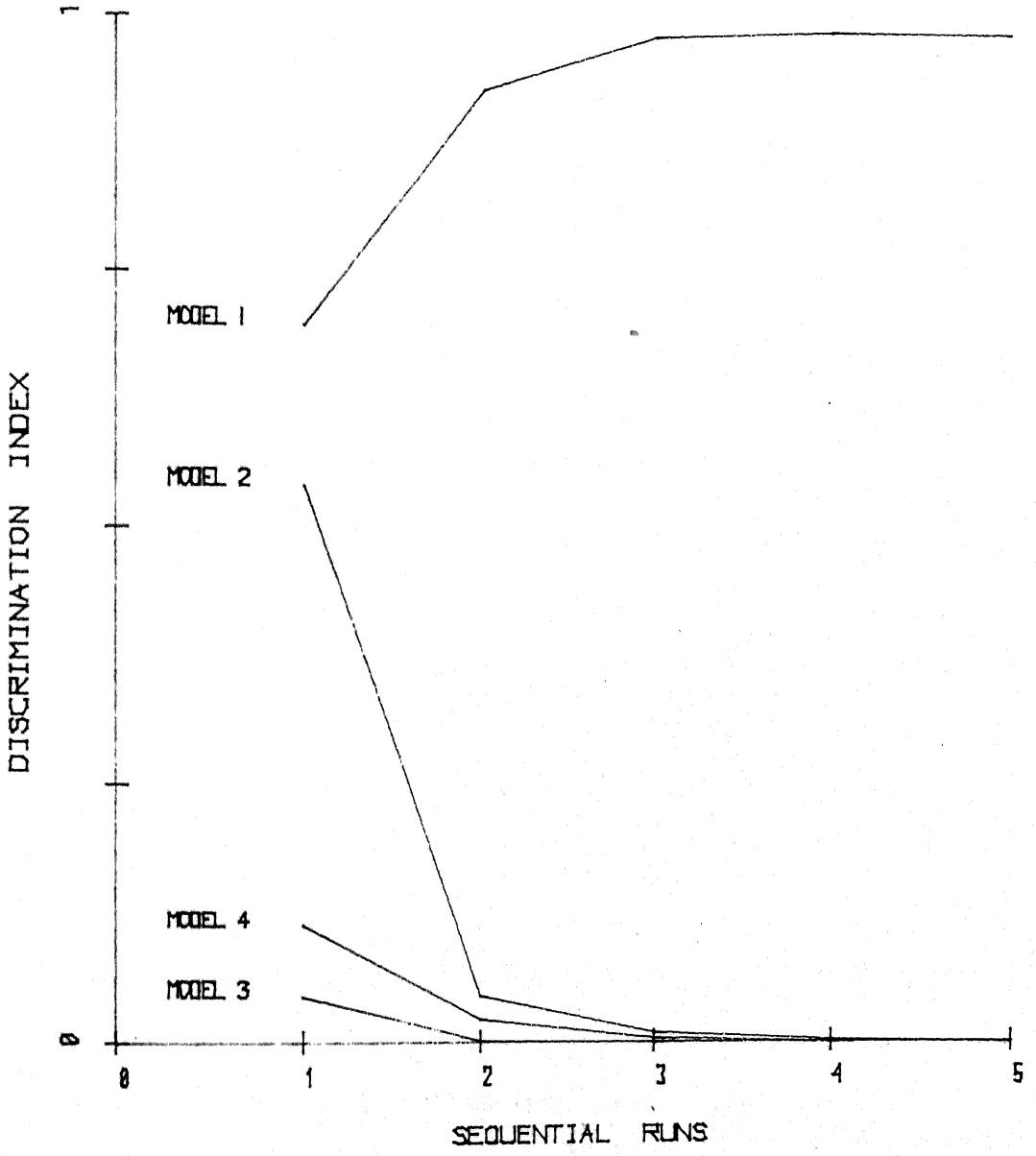

FIGURE 6.2.1 Status of models as determined by discrimination index; Linear univariate models. True model: Model 3.

Fortunately, in this problem it rose to 0.90 and still higher to 0.97 when according to Box and Hill (1967) one could take a decision that $M^{(3)}$ was the best model. Nevertheless, because of the fluctuating nature of their discrimination criterion it was not safe to declare $M^{(3)}$ as the best model just on the basis of the initial set of observations or even after 3 more points had been designed. It actually took them 9 runs, in total, to take the final decision. On the other hand the behaviour of the discrimination index being more stable it is safe to take the decision even on the basis of 5 preliminary observations, if the need were.

Furthermore, it may be noted that the Box-Hill method puts more stress on bad models. For example, at the first instance itself the probabilities of models 1 and 2 are rendered very low; being, 0.00, 0.01, respectively. Even at later stages the fall in the probabilities of these models can be seen to be much faster as compared to the rise in the probabilities of better models. On the other hand, the present procedure gives more importance to the better models as can be observed through the values, 0.6981, 0.1687, 0.0197, 0.1135 of the discrimination index to start with and in the later stages.

Another point which adds to the usefulness of the procedure proposed here is concerned with the adequacy of the selected model. At the 4th sequential stage where we decide to stop we notice that the value of the statistic, V for model 3; is

as low as 0.0002. This value of V is insignificant as compared to the 5% point, 3.841, of the Chi-square distribution with 1 degree of freedom. We, therefore, conclude that model $M^{(3)}$ is not only the best model, as decided earlier, but also an adequate model for the data simulated from the normal distribution with expectation given through the polynomial, $1+\xi+\xi^2$, and variance, 1.0. To be more specific, the model,

$$Y = 0.7027 + 0.7601\,\xi + 1.0740\,\xi^2,$$

as seen at the termination of our procedure is a ready-to-use model in the given set-up.

It has been observed by many investigators that the points designed for discrimination through some of the procedures are not good for estimation. Therefore, the parameters of the model selected through these methods normally require polishing before one can put it to use. It can be actually seen through Table 6.2.2 that the Box-Hill procedure in this example always results into 0.0 as the optimal value of ξ_6. On the other hand, the present procedure generates points such as 1.6 and 4.0, when the design space has been considered to be $0.0 \leq \xi \leq 4.0$. In fact, while working on the real life systems, it is recommended that occassionally experiments be chosen in the interior of the region, even when not prescribed by the design criterion.

Lastly, we discuss the role of weights, $w_{u,v;k}$, in designing additional experiments for the problem under consideration. It may be observed through Table 6.2.2 or in Figure 6.2.1 that the pair, $(M^{(2)}, M^{(4)})$, is found to be the closest.

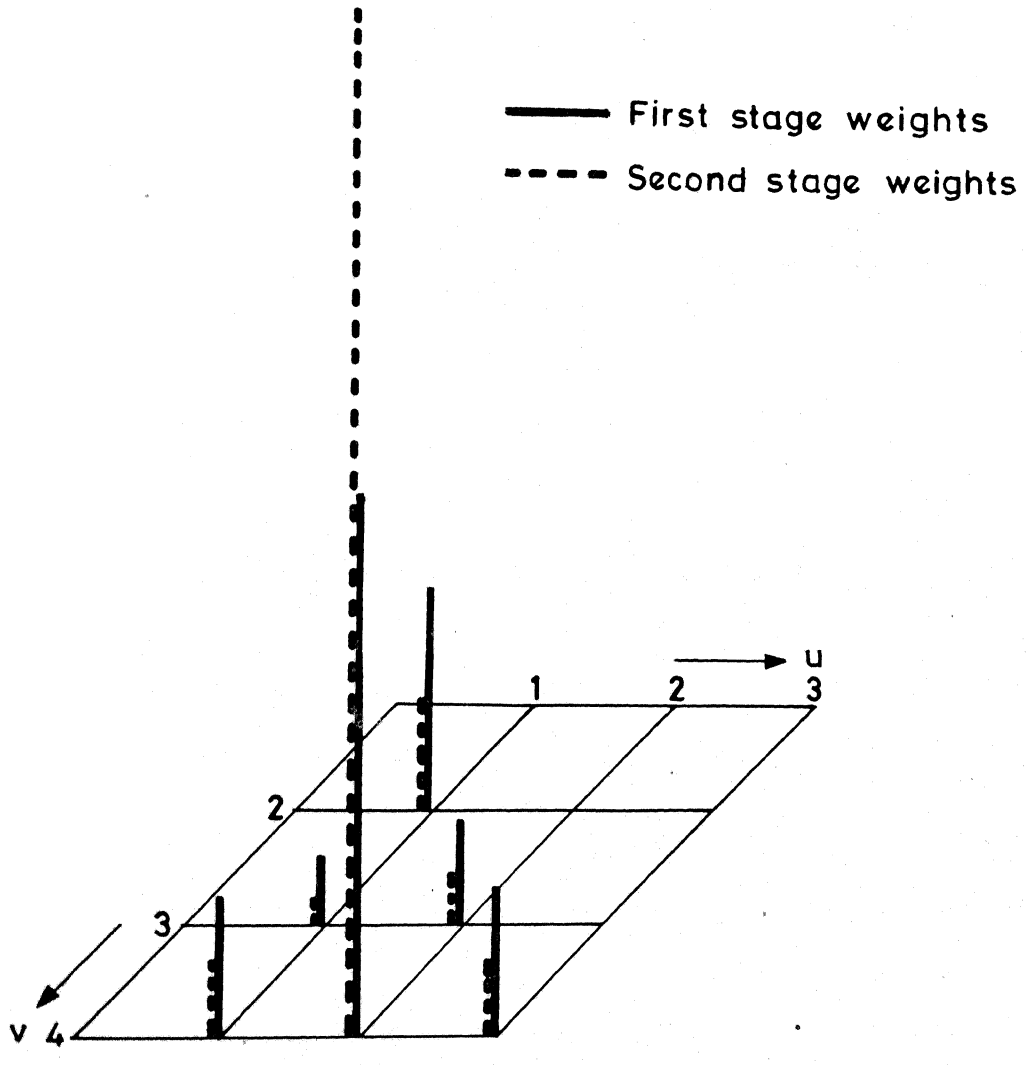

FIGURE 6.2.2 Change in weights, $w_{u,v;k}$, from first to second stage; linear univariate models.

Table 6.2.3
Weights used in the criterion function ϕ at different stages: discrimination among polynomial models

Sequential stage k	Model v	Model u		
		1	2	3
1	2	0.172998		
	3	0.020281	0.083937	
	4	0.116371	0.481631	0.124782
2	2	0.078665		
	3	0.000972	0.020478	
	4	0.038918	0.819574	0.041392
3	2	0.023600		
	3	0.000032	0.004063	
	4	0.074030	0.951948	0.012954
4	2	0.007617		
	3	0.000001	0.001005	
	4	0.001114	0.983388	0.006874
5	2	0.002173		
	3	0.000000	0.000261	
	4	0.000150	0.993641	0.003775

at the end of the fifth run, receives most of the attention, i.e., 48%, in designing the 6th experiment. Similarly, the pair $(M^{(1)},M^{(3)})$ consisting of the farthest models have been given the least importance through a weight of 0.020281, only. The other pairs, too are given their due through the weights, 0.172998 $(M^{(1)},M^{(2)})$, 0.116371 $(M^{(1)},M^{(4)})$, 0.083137 $(M^{(2)},M^{(3)})$, and 0.124782 $(M^{(3)},M^{(4)})$ in designing the sixth setting of ξ. The values of weights assigned to different pairs of models at various sequential stages are shown in Table 6.2.3, where it can be easily seen that throughout the design process the pair $(M^{(3)},M^{(4)})$ kept on receiving the highest importance, while the pair $(M^{(1)},M^{(3)})$ has always been given the least weight as it never required further divergence. This way a look at the values of the discrimination index at any stage, in Table 6.2.2, clearly indicates that the proposed weights appropriately decide the role of a pair of models in designing an additional experiment.

Finally, in order to examine the proposed discrimination criterion for its immunity to sampling fluctuations and attach as well a confidence level with our contention, we have simulated 500 samples of the response using random samples of the values of ξ, uniformly distributed over its operability region, $\{0.0 \leq \xi \leq 4.0\}$. We find that 489 times the criterion picks up the correct model; although, sometimes discrimination based on a set of initial data is not found to be very sharp. As regards the remaining 11 samples, the distinction among models

is not seen to be clear, although none of these have shown $M^{(3)}$ to be a bad model.

6.2.2 Example : Nonlinear Models : Unknown Homogeneous Error Variances

The following example has been considered by Buzzi and Forzatti (1983) for demonstrating the implementation of their discrimination scheme and for its comparison with the Box-Hill procedure. We include it here for comparison of our method with that of Box and Hill (1967) and with the one given by Buzzi and Forzatti (1983). In this example we, therefore, consider the same set of five kinetic models; namely,

$$M^{(1)} : \eta^{(1)} = \frac{\xi_1\xi_2^2 - (\xi_3/\theta_1^{(1)})}{(\theta_2^{(1)} + \theta_3^{(2)}\xi_1 + \theta_4^{(1)}\xi_2 + \theta_5^{(1)}\xi_3)^2} ,$$

$$M^{(2)} : \eta^{(2)} = \frac{\xi_1\xi_2^2 - (\xi_3/\theta_1^{(2)})}{(\theta_2^{(2)} + \theta_3^{(2)}\xi_1 + \theta_4^{(2)}\xi_2 + \theta_5^{(2)}\xi_1\xi_2)} ,$$

$$M^{(3)} : \eta^{(3)} = \frac{\xi_1\xi_2^2 - (\xi_3/\theta_1^{(3)})}{\{\theta_2^{(3)} + \theta_3^{(3)}\xi_3 + \theta_4^{(3)}\xi_2 + (\theta_5^{(3)}\xi_3/\xi_2)\}\xi_2^2} ,$$

$$M^{(4)} : \eta^{(4)} = \frac{\xi_1\xi_2^2 - (\xi_3/\theta_1^{(4)})}{\{\theta_2^{(4)} + \theta_3^{(4)}\xi_1 + (\theta_4^{(4)}\xi_3/\xi_2) + \theta_5^{(4)}\xi_3\}^2\xi_2} ,$$

$$M^{(5)} : \eta^{(5)} = \frac{\xi_1\xi_2^2 - (\xi_3/\theta_1^{(5)})}{(\theta_2^{(5)} + \theta_3^{(5)}\xi_2 + \theta_4^{(5)}\xi_1\xi_2 + \theta_5^{(5)}\xi_3)^2} .$$

These models have, in fact, been originally proposed for the process of synthesis of menthol from carbon monoxide and

hydrogen. We, however, like Buzzi and Forzatti, carry out a Monte Carlo study and simulate experiments through the model

$$y = \frac{\xi_1 \xi_2^2 - 58824.0\, \xi_2}{(1704.0 + 4.25\, \xi_2 + 0.241\, \xi_1 \xi_2 + 444.6\, \xi_3)^2} + \varepsilon, \qquad (6.2.2)$$

with ε as a pseudorandom number from $N(0, 4.0 \times 10^{-6})$. While designing the optimal settings of ξ_1, ξ_2, and ξ_3, we have likewise restricted these variables within their respective operability regions; namely, $\{15 \leq \xi_1 \leq 25\}$, $\{200 \leq \xi_2 \leq 250\}$, and $\{5 \leq \xi_3 \leq 10\}$. The corresponding observations are simulated through the model (6.2.2). So far as the initial set of observations is concerned we have used the data simulated by Buzzi and Forzatti (1983), based on a 3^2 factorial design of ξ_1, ξ_2, ξ_3. On the basis of these 8 simulated observations the Box-Hill method produced posterior probabilities of five models as, 0.3000, 0.0007, 0.0400, 0.3290, and 0.3290, which hardly show any distinction among $M^{(1)}, M^{(4)}$, and $M^{(5)}$. Even the estimates, 4.83×10^{-6}, 16.8×10^{-6}, 8.78×10^{-6}, 4.65×10^{-6}, and 4.65×10^{-6}, of $S_8^{(u)}$'s in the method of Buzzi and Forzatti, which are claimed to, atleast, screen out bad models, do not seem to do the job. The same set of data when used in (2.4.1) gives the values of the discrimination index as 0.0501, 0.7025, 0.1468, 0.1001, 0.0005, for the models $M^{(1)}, M^{(2)}, M^{(3)}, M^{(4)}$, and $M^{(5)}$, respectively. These values clearly indicate the superiority of $M^{(5)}$, i.e., the model which generated the data and the

inferiority of $M^{(2)}$. At this stage, both the posterior probability and the criterion of Buzzi and Forzatti no doubt show that $M^{(2)}$ is a bad model but are not even slightly indicative of the reality which we are actually interested in revealing. Thus, one can see more discriminating power of the discrimination index as compared to the other two criteria, when only initial set of data have been used.

The inability of the other two discrimination criteria in choosing the best model at this stage necessitated the designing of additional points as was done through their respective design criteria. The results obtained are presented in Table 6.2.4, where the values bracketed as ( ) are obtained through the Box-Hill procedure and those bracketed as [ ] come through the discrimination scheme of Buzzi and Forzatti. It may be noted from this table that the values of $P_k^{(2)}$ and $P_k^{(3)}$ keep decreasing throughout the sequential process and finally go as low as 0.0, thereby showing that $M^{(2)}$ and $M^{(3)}$ are poor models. On the other hand, $M^{(1)}$ remains the least favoured model upto the 11th run only, after which it starts showing up as the best model; the preference shifts from $M^{(4)}$ and $M^{(5)}$ to $M^{(1)}$. Whatsoever, the Box-Hill procedure could not pick up the true model. Instead it declared $M^{(1)}$ as the most probable model. Similarly upto a considerable number of runs the values of $S_k^{(u)}$, too, do not seem to be indicative of the fact that $M^{(5)}$ is the correct model. For example, at a stage as late as 30th the values of $S_{30}^{(u)}$ for different models are :

Table 6.2.4

Sequential discrimination among nonlinear models; the present, the Box-Hill, and the Buzzi et al. procedure

(True model : $\underline{y}(5)$)

Run k	Input Variables ξ_{1k}	ξ_{2k}	ξ_{3k}	Response y_k	Discrimination criteria $D_k^{(1)}$ ($P_k^{(1)}$) [$S_k^{(1)}$]	$D_k^{(2)}$ ($P_k^{(2)}$) [$S_k^{(2)}$]	$D_k^{(3)}$ ($P_k^{(3)}$) [$S_k^{(3)}$]	$D_k^{(4)}$ ($P_k^{(4)}$) [$S_k^{(4)}$]	$D_k^{(5)}$ ($P_k^{(5)}$) [$S_k^{(5)}$]
1	17	210	6						
2	23	210	6						
3	17	240	6						
4	23	240	6						
5	17	210	9	0.104					
6	23	210	9	1.041					
7	17	240	9	0.764					
8	23	240	9	1.097					
					0.0501	0.7025	0.1468	0.1002	0.0004
					(0.3000)	(0.0007)	(0.0004)	(0.3290)	(0.3290)
					[4.83]	[16.8]	[8.78]	[4.65]	[4.65]
9	25	202	5	2.202	0.0037	0.9447	0.0303	0.0213	0.0000
	(25)	(230)	(5)	(2.530)	(0.2930)	(0.0000)	(0.0003)	(0.3520)	(0.3530)
	[25]	[230]	[5]	[2.530]	[3.87]	[18.00]	[8.26]	[3.73]	[3.72]
10	24	250	5	2.910	0.0002	0.9944	0.0040	0.0014	0.0000
	(15)	(250)	(5)	(1.500)	(0.3220)	(0.0000)	(0.0007)	(0.3220)	(0.3550)
	[15]	[250]	[5]	[1.500]	[5.11]	[15.00]	[6.88]	[5.35]	[5.32]
11	24	206	5	2.0101	1.0E-6	0.9994	0.0005	0.0002	0.0000
	(25)	(250)	(10)	(1.030)	(0.4080)	(0.0000)	(0.0001)	(0.2690)	(0.0000)
	[25]	[250]	[10]	[1.030]	[4.69]	[19.5]	[6.81]	[5.16]	[5.06]

(i) Values brackted as [] are to be multiplied by 1.0E-6,

(ii) 1.0E-k stands for 10^{-k} .

4.16×10^{-6}, 15.4×10^{-6}, 4.46×10^{-6}, 4.14×10^{-6}, and 4.21×10^{-6}, and the prescribed test on these values do not show any distinction among the closer models, i.e., $M^{(1)}$, $M^{(4)}$, and $M^{(5)}$. By using the procedure of Buzzi and Forzatti we only end up with the conclusion that the models, $M^{(1)}$, $M^{(3)}$, $M^{(4)}$, and $M^{(5)}$ can all be considered to be appropriate for the simulated process; according to them these models are equivalent on statistical grounds. This conclusion which has been made possible after 30 experiments could have been drawn through the discrimination index at the initial stage itself. In fact, from the preliminary data only the index could extract much information : it not only showed the equivalence of $M^{(1)}$, $M^{(3)}$, and $M^{(4)}$ but also established the truth that $M^{(5)}$ is the model which seems to have generated the data.

It is important to compare the three procedures at the 11th run, for it is at this stage that the present procedure may be terminated. A close look at the values of the three discrimination criteria in Table 6.2.4 show that at this stage when the Box-Hill procedure just starts favouring rather model, $M^{(1)}$, and the method of Buzzi and Forzatti has just screened out one model; being $M^{(2)}$, the discrimination index has already shown enough evidence in favour of $M^{(5)}$ as well as given a clear indication as to which of the rival models are close to $M^{(5)}$. This exhibits the potential power of the procedure proposed here.

It may further be noted that because of $P_{30}^{(2)}$ being zero and $S_{30}^{(2)}$ $(= 15.4 \times 10^{-6})$ being significant as compared to the estimate of the experimental error variance, σ^2 $(= 11.3 \times 10^{-6})$ the other two procedures have been quite decisive in rejecting $M^{(2)}$; both the procedures seem to put more stress on the identification of the bad model rather than distinguishing $M^{(5)}$ from $M^{(1)}$ and $M^{(4)}$. However, the results obtained through the two methods lead to different conclusions; whereas one prefers model $M^{(1)}$ over all other models, the other considers it to be the same as $M^{(4)}$ and $M^{(5)}$. But, the fact that $M^{(5)}$ is the model which have been generating the data ought not be ignored.

So far as the performance of the proposed design and discrimination criteria in identifying the true model is concerned 3 more additional experiments have been designed, although picture had been made quite clear the the 8th run itself. The results obtained are shown in Table 6.2.4. It can be seen from this table that the decision on the best model remains invariant throughout the sequential process. One can declare $M^{(5)}$ as the true model at any stage; in fact, more and more decisively as one moves from one stage to the next. Furthermore, it is easy to see that barring $M^{(2)}$, the other models, i.e., $M^{(1)}$, $M^{(3)}$, and $M^{(4)}$ are close to $M^{(5)}$.

Finally, when we stop our procedure at the 11th run according to the proposed stopping rule and declare $M^{(5)}$ as the

best model with $D_{10}^{(5)} = 0.9 \times 10^{-6}$ we also establish that this model is adequate for the simulated process. In fact, the value of the statistic V for $M^{(5)}$, being 0.00001, is much less as compared to the 5% point, $(\chi^2_{0.05;1} = 3.842)$, of the Chi square distribution with one degree of freedom. This enables us to accept the hypothesis on the adequacy of $M^{(5)}$. Similarly, when the values, 0.0118, 0.0910, 0.9630, of the statistic V for models $M^{(1)}$, $M^{(3)}$, and $M^{(4)}$, respectively, are compared with the above tabulated value we conclude that on statistical grounds these models can also be considered to be good. The use of the χ^2 approximation in testing the above hypotheses is justified by the small value of w_2, being 0.00023. Thus the ready-to-use form of the model $M^{(5)}$ may be specified as

$$E(Y) = \frac{\{\xi_1 \xi_2^2 - (\xi_3/0.0002)\}}{\{1704.0015 + 4.25\ \xi_2 + 0.2408\ \xi_1\xi_2 + 443.9291\ \xi_3\}^2} .$$

In addition, on the basis of the discrimination index we also conclude that $M^{(1)}$ is the closest model to $M^{(5)}$.

Figure 6.2.3 is given here to present a picture of the progress in discrimination achieved in this problem through the proposed procedure. It may be observed from this figure that the graph G5 always remains below the graphs, G1 , G3 , G4 , and G2 . This establishes the superiority of $M^{(5)}$ over other residual models throughout the sequential procedure. However, the falling trends of G1 , G3 , and G4 also maintain that $M^{(1)}$, $M^{(3)}$, and $M^{(4)}$, too, may be

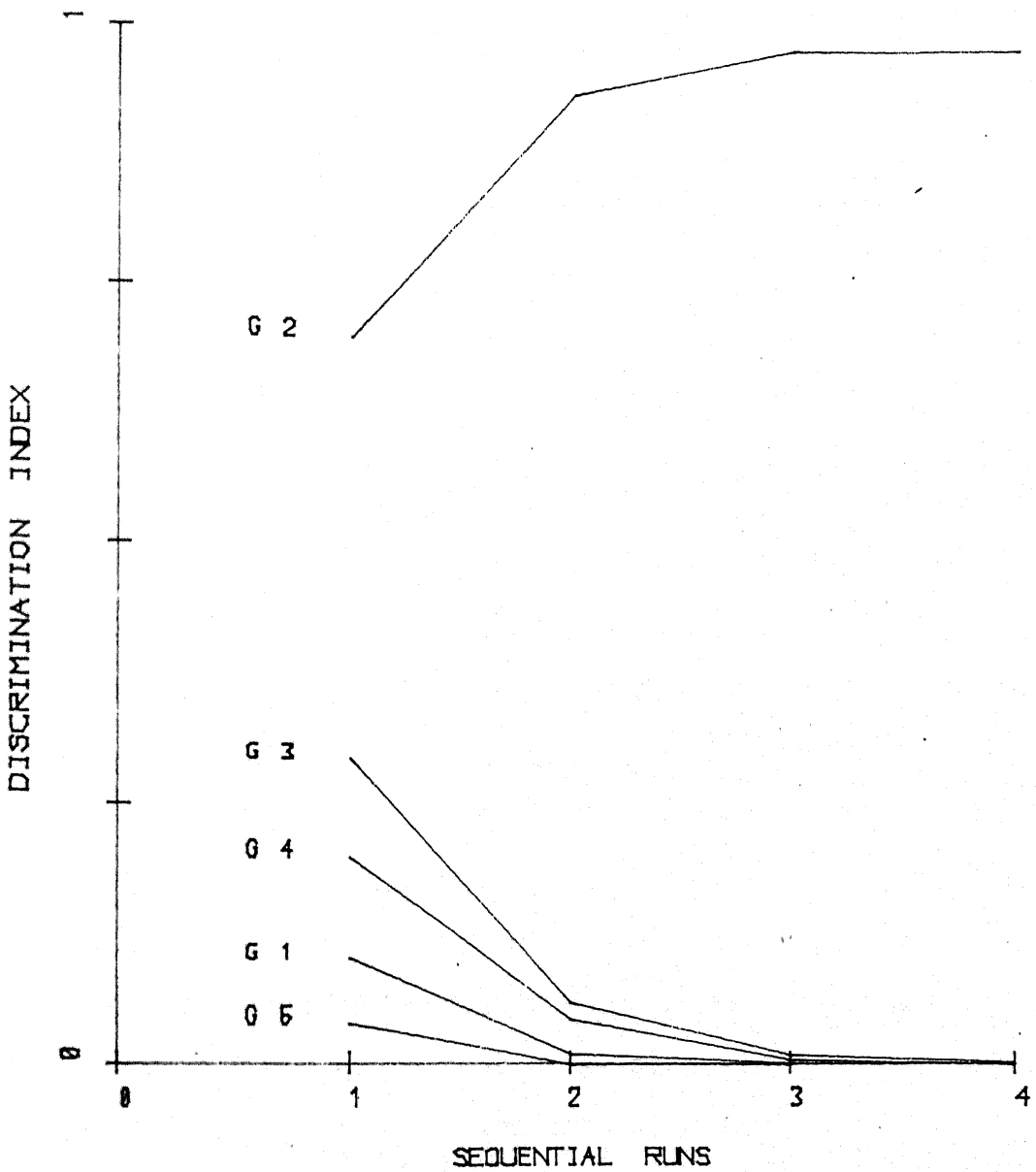

FIGURE 6.2.3 Progress in discrimination through the proposed procedure; nonlinear univariate models. True model : Model 5 .

considered to be good models. On the other hand, G2 keeps rising consistently. This confirms that $M^{(2)}$ is a bad model. These are some of facts which have already been proved through the proposed adequacy test.

As far as the role of the weights, $w_{u,v;k}$, in designing new experiments is concerned it can be seen through Table 6.2.5 that the weights attached with different pairs are determined according to the position of a particular pair relative to all other pairs. For example, in Table 6.2.4 we find that the farthest models are $M^{(2)}$ and $M^{(5)}$ while $M^{(3)}$ and $M^{(4)}$ make a pair with the closest models. Accordingly, while designing the 9th experiment the pair $(M^{(2)}, M^{(5)})$ receives the least weight, i.e., 0.000269 , while the pair $(M^{(3)}, M^{(4)})$ is attached with the maximum weight of 0.347799 as can be seen through Table 6.2.5. The next closest pair being $(M^{(1)}, M^{(4)})$ is given a weight of 0.255276 . Similarly, the other pairs, too, are attached with due weights. Figure 6.2.4 gives a pictorial view of the allocation of weights for designing the 9th experiment. It can be seen through a combined study of Tables 6.2.4 and 6.2.5 that at every stage heavy weights are assigned to models which are yet to be diverged.

6.2.3 Example : Nonlinear Models: Unknown, Homogeneous Error Variances

Consider the problem of discriminating among the following four models proposed for a chemical reaction .

Table 6.2.5
Weights for the criterion function ϕ at different stages discrimination among nonlinear univariate models

Sequential stage k	Model v	Model u			
		1	2	3	4
1	2	0.036390			
	3	0.174152	0.106522		
	4	0.255276	0.072670	0.347799	
	5	0.003768	0.000269	0.001287	0.001887
2	2	0.003732			
	3	0.116208	0.030357		
	4	0.165820	0.021274	0.662385	
	5	0.000171	0.000001	0.000021	0.000030
3	2	0.000356			
	3	0.089144	0.007287		
	4	0.250230	0.002569	0.650377	
	5	0.000009	0.000000	0.000000	0.000001

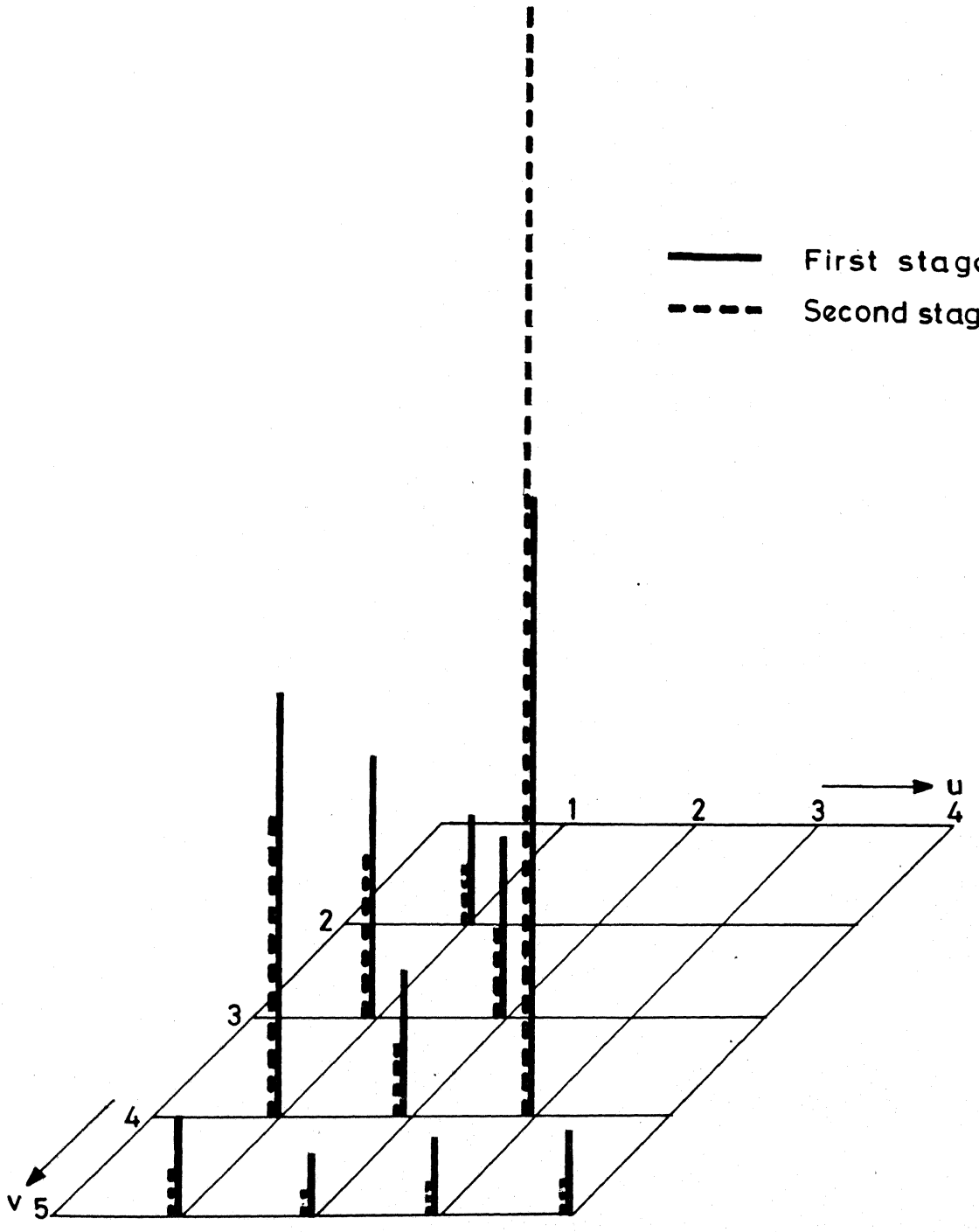

FIGURE 6.2.4 Change in weights, $w_{u,v;k}$, from first to secon stage; nonlinear univariate models.

$$M^{(1)} : \eta^{(1)} = \exp\{-\xi_1 \exp(\theta_1^{(1)} - \theta_2^{(1)}\xi_2)\} ,$$

$$M^{(2)} : \eta^{(2)} = \{1 + \xi_1 \exp(\theta_1^{(2)} - \theta_2^{(2)}\xi_2)\}^{-1} ,$$

$$M^{(3)} : \eta^{(3)} = \{1 + 2\xi_1 \exp(\theta_1^{(3)} - \theta_2^{(3)}\xi_2)\}^{-1/2} ,$$

$$M^{(4)} : \eta^{(4)} = \{1 + 3\xi_1 \exp(\theta_1^{(4)} - \theta_2^{(4)}\xi_2)\}^{-1/3} .$$

It is the amount of reactant which is supposed to be measured in the reaction under consideration. Besides, ξ_1 is the time, t, and ξ_2 [$= (\frac{1}{T} - \frac{1}{525})$] is the scaled inverse absolute temperature, with T representing the temperature. In actual experimentation the experimenter would probably like to control the input variables t and T over the grid {t = 25(25)150, T = 450(25)600}. In the simulation of the problem, as we carry out here, the same grid is used for the selection of time and temperature for generating observations. Further, we assume that $M^{(2)}$ is the true model and accordingly simulate the amount of reactant, y using the equation

$$M^{(0)} : y = [1 + \xi_1 \exp\{-(3.53235 + 5000.0\,\xi_2)\}]^{-1} + \varepsilon, \qquad (6.2.3)$$

with ε being a pseudorandom number from $N_1(0, 0.0025)$. This set-up is the same as used by Hill et al. (1968). To start with, we use the data simulated by them corresponding to the settings of t and T according to a 2^2 factorial design. After securing estimates of the parameters in all the four competing models through $\mathcal{C}_4$ the values of the discrimination index for the models are found to be 0.4310, 0.0071, 0.2295, and 0.3324. Thus

the preliminary data shows large differences between the value of $D_4^{(2)}$ and that of $D_4^{(1)}$, $D_4^{(3)}$, and $D_4^{(4)}$. This is indicative of the superiority of $M^{(1)}$. On the other hand, with the same preliminary data the posterior probability formula of Box and Hill gives 0.0060, 0.4335, 0.4087, 0.1518 as the respective probabilities of the four models.

When another experimental setting; namely (150, 525) is decided through a search of the maximum of ϕ over the prescribed grid, utilizing h_2 of (3.3.32), the values of the index for $M^{(1)}$ and $M^{(4)}$ rise to 0.5108 and 0.3397, respectively; $D_5^{(2)}$ drops to 0.0003; and $D_5^{(3)}$ decreases to 0.1493. It clearly shows that $M^{(2)}$ is the best model for the data simulated through (6.2.3) and that $M^{(3)}$ is its closest rival. The Box-Hill criterion suggests different settings of t and T, i.e., (150, 550). The posterior probabilities which are obtained as result do not seem to differentiate reasonably well between $M^{(2)}$ and $M^{(3)}$; $P_5^{(2)}$ and $P_5^{(3)}$ being of the orders 0.5580 and 0.3740, respectively; $M^{(1)}$ no doubt has been clearly shown to be a poor model with $P_5^{(1)} = 0.0004$. The progress in discrimination through the two procedures can be studied through Table 6.2.6, where the bracketed values come through the Box Hill procedure used by Hill et al. (1968). Figures 6.2.5 and 6.2.6 show the trend of discrimination through the two procedures from one stage to another. In fact, these figures give a pictorial view of the changes in the relative status of the competing models as decided by the two discrimination criteria. The procedure proposed here

Table 6.2.6
Sequential Discrimination among nonlinear univariate models: the present and Box-Hill procedures
(True model: $M^{(2)}$)

Run k	Input Variables ξ_{1k}	ξ_{2k}	Response y_k	Discrimination Criteria $D_k^{(1)}$ ($P_k^{(1)}$)	$D_k^{(2)}$ ($P_k^{(2)}$)	$D_k^{(3)}$ ($P_k^{(3)}$)	$D_k^{(4)}$ ($P_k^{(4)}$)
1	25	575	0.3961				
2	25	475	0 7232				
3	125	475	0.4215				
4	125	575	0.1297	0.4310	0.0071	0.2295	0.3324
	(125)	(575)	0.1297	(0.0060)	(0.4335)	(0.4087)	(0.1518)
5	150	525	0.1875	0.5108	0.0003	0.1493	0.3397
	(150)	(550)	0.1504	(0.0004)	(0.5580)	(0.3740)	(0.0676)
6	25	525	0.5774	0.5433	0.0000	0.0983	0.3584
	(25)	(525)	0.5565	(0.0001)	(0.6278)	(0.3446)	(0.0275)
7	150	525	0.1912	0.5702	0.0000	0.0637	0.3661
				the present procedure is stopped			
	(150)	(550)	0.1558	(0.0000)	(0.6865)	(0.3020)	(0.0115)
8	(150)	(550)	0.0671	(0.0000)	(0.9424)	(0.0570)	(0.0005)
9	(25)	(525)	0.6169	(0.0000)	(0.9895)	(0.0105)	(0.0000)
10	(150)	(525)	0.1427	(0.0000)	(0.9978)	(0.0022)	(0.0000)
11	(25)	(525)	0.5979	(0.0000)	(0.9993)	(0.0007)	(0.0000)
12	(150)	(525)	0.1717	(0.0000)	(0.9997)	(0.0003)	(0.0000)
13	(150)	(525)	0.2419	(0.0000)	(0.9995)	(0.0005)	(0.0000)
14	(25)	(525)	0.5441	(0.0000)	(0.9994)	(0.0006)	(0.0000)
15	(150)	(525)	0.1977	(0.0000)	(0.9996)	(0.0004)	(0.0000)

suggests termination at the third sequential stage with the decision that $M^{(2)}$ is the adequate model for the system simulated through $M^{(0)}$ of equation (6.2.3). On the other hand, Hill et al. could stop the Box Hill procedure after designing 11 additional experiments, when the posterior probability of $M^{(2)}$ rose to 0.9996. At the same time it is worth nothing that through the Box-Hill procedure one could declare $M^{(1)}$ as the worst model much earlier i.e. after one had designed 2 additional experiments only. An important point as can be brought out through this observation is that in the Box-Hill method more emphasis is on bad models rather than on good ones. It can be seen through Table 6.2.6 that the posterior probability shows large differences in the best and the worst models, while discrimination between close rivals, $M^{(2)}$ and $M^{(3)}$, is much slower. This is not the case with the present method, which on the other hand puts more stress on good models and discriminates sharply between the closest rivals, as can be seen through Table 6.2.6. This is the effect of the weights being used in the criterion function φ; the closer models are always given more weightage in designing experiments than the ones which are already farther at the previous stage. Finally, we notice in Table 6.2.6 that the points obtained through the two procedures are different, though both arrive at the same conclusion; namely model $M^{(2)}$ is the correct model.

6.3 DISCRIMINATION AMONG MULTIVARIATE MODELS

6.3.1 Example : Nonlinear Models; Unknown Equal Error Covariance Matrices

Buzzi et al. (1984) have simulated discrimination among the following four bivariate chemical kinetic models through their procedure.

$$M^{(1)}: \quad \eta_1^{(1)} = \theta_1^{(1)}\xi_1\xi_2/(1 + \theta_3^{(1)}\xi_1 + \theta_4^{(1)}\xi_2)$$

$$\eta_2^{(1)} = \theta_2^{(1)}\xi_1\xi_2/(1 + \theta_3^{(1)}\xi_1 + \theta_4^{(1)}\xi_2) \quad ,$$

$$M^{(2)}: \quad \eta_1^{(2)} = \theta_1^{(2)}\xi_1\xi_2 / (1 + \theta_3^{(2)}\xi_1 + \theta_4^{(2)}\xi_2)^2$$

$$\eta_2^{(2)} = \theta_2^{(2)}\xi_1\xi_2 / (1 + \theta_3^{(2)}\xi_1)^2 \quad ,$$

$$M^{(3)}: \quad \eta_1^{(3)} = \theta_1^{(3)}\xi_1\xi_2 / (1 + \theta_3^{(3)}\xi_2)^2$$

$$\eta_2^{(3)} = \theta_2^{(3)}\xi_1\xi_2 / (1 + \theta_4^{(3)}\xi_1)^2 \quad ,$$

$$M^{(4)}: \quad \eta_1^{(4)} = \theta_1^{(4)}\xi_1\xi_2 / (1 + \theta_3^{(4)}\xi_1 + \theta_4^{(4)}\xi_2)$$

$$\eta_2^{(4)} = \theta_2^{(4)}\xi_1\xi_2 / (1 + \theta_3^{(4)}\xi_1).$$

We take up discrimination among these models in order to see the potential of our procedure in identifying the correct model and compare its performance with the method of Buzzi et al. So far as the data are concerned they have used $M^{(1)}$ as the true model. In this connection, it may, be pointed out that the values, 0.1, 0.01, 0.1, and 0.01 of the parameters of this model

as reported in their paper,do not seem to be the ones which have been actually used for simulation of responses. Assuming that the data reported are correct,the values for the parameters of $M^{(1)}$ used in simulation should rather be 0.01, 0.001, 0.01, and 0.001. To start with we use the data of Buzzi et al. (1984) in our discrimination criterion - the Discrimination Index and find that $M^{(1)}$ with 0.0134 as the value of this index is closest to the true model. On the other hand, the highest value 0.4040 of $D_9^{(3)}$ is indicative of the worst performance of $M^{(3)}$. Besides, the values, 0.3088 and 0.2738 of $D_9^{(2)}$ and $D_9^{(4)}$, respectively show the closeness between $M^{(2)}$ and $M^{(4)}$, although as compared to $M^{(1)}$ these prove to be inferior models.

In order to make the distinction more clear we design some additional points, using the assumption that the covariance matrix of errors, under the four models, is same, though unknown. To that purpose the independent variables (ξ_1,ξ_2) are constrained to lie in the interval $\{5.0 \leq \xi_1, \xi_2 \leq 55.0\}$ and the criterion $\phi(\underset{\sim}{\xi}_{9+1})$, utilizing h_2 of (4.3.28) is employed. So far as the weights for the criterion function are concerned, models $M^{(2)}$ and $M^{(4)}$ being closest at the 9th run are given the highest weightage, 0.361349, while $M^{(1)}$ and $M^{(3)}$, being the farthest, receive the least, i.e., 0.013503. The criterion yields (54.0, 28.0) as the new optimal point. For the purpose of simulating the values of the responses y_1,y_2 we use the model

$$M^{(0)}: \quad y_1 = (0.01\,\xi_1\xi_2)/(1 + 0.001\,\xi_1 + 0.01\,\xi_2) + \varepsilon_1$$

$$y_2 = (0.001\xi_1\xi_2)/(1 + 0.001\,\xi_1 + 0.01\,\xi_2) + \varepsilon_2 \qquad (6.3.1)$$

as the true model with $(\varepsilon_1,\varepsilon_2)$ as pseudorandom numbers from $N_2(0,\Sigma)$,where Σ = diag $\{0.35,0.0023\}$. This way,corresponding to the new point we obtain (9.66,0.61)as the values of the two responses. The addition of these values to the sample of 9, further results into a considerable decrease in the value of $D_{10}^{(1)}$; dropping to 0.0004. On the other hand $D_{10}^{(3)}$ rises to 0.4962. The models $M^{(2)}$ and $M^{(3)}$, too, look more apart from each other as can be seen from Table 6.3.1. Since the termination condition (2.6.1) is not satisfied we continue the sequential procedure, although model $M^{(1)}$ seem to have been considerably diverged from $M^{(2)}$, $M^{(3)}$, and $M^{(4)}$. It is for this reason that the pairs $(M^{(1)}, M^{(2)})$, $(M^{(1)},M^{(3)})$, and $(M^{(1)},M^{(4)})$ receive low weights of the orders of 0.0007, 0.0004, and 0.0011, respectively for the next design point. This time the point (17.0, 45.0) is designed, resulting into (6.30, 0.52) as the values of the simulated responses. At this stage there is a further drop in $D_{11}^{(1)}$; i.e., to 0.00001, which clearly indicates that the procedure has identified the true model. However, it is only at the third sequential stage that the procedure is stopped according to the termination criterion. By this time the value of the discrimination index for $M^{(1)}$ has gone as low as 0.1752×10^{-6} and is decisive of the fact that this model is the one which generated the data. We also find at this

stage that $M^{(4)}$ with $D_{12}^{(4)} = 0.0916$ is closest to $M^{(1)}$ and that $M^{(3)}$ with $D_{12}^{(3)} = 0.6790$ may be considered to be a bad model. The progress in discrimination in the sequential procedure can be seen through Table 6.3.1. Besides Figure 6.3.1 is presented to see the trend of discrimination. Table 6.3.2 shows weights used at different stages and Figure 6.3.2 presents a picture of the change in weights according as the relative position of the models changes from the first stage to the second.

Finally, we test the adequacy of the selected model, i.e.,

$$M^* : y_1 = \frac{0.0099\,\xi_1\xi_2}{(1 + 0.0991\,\xi_1 + 0.000854\,\xi_2)}$$

$$y_2 = \frac{0.00098\,\xi_1\xi_2}{(1 + 0.991\,\xi_1 + 0.000854\,\xi_2)} .$$

Using the test proposed for this purpose we find that the value of the statistic V (= 0.005) is much smaller than the corresponding 5% point of the Chi-square distribution ($\chi^2_{3;0.05} = 7.815$), thereby establishing that M^* is an adequate model for the simulated chemical process.

Table 6.3.3 presents the results of discrimination, simulated by Buzzi et al. (1984) among the same set of models. Since the discrimination, according to their procedure, is assessed through the model adequacy, the initial set of data do not show any distinction among the models. In fact, the values, 15.1. 14.7, 24.3, and 14.4 of the statistic

Table 6.3.1
Sequential discrimination among nonlinear bivariate models : the present procedure
(True model : $M^{(1)}$)

Run k	Input variables ξ_{1k}	ξ_{2k}	Responses y_{1k}	y_{2k}	Discrimination index $D_k^{(1)}$	$D_k^{(2)}$	$D_k^{(3)}$	$D_k^{(4)}$
1	20	20	3.61	0.53				
2	30	20	5.42	0.44				
3	20	30	5.00	0.64				
4	30	30	7.50	0.66				
5	25	25	5.73	0.55				
6	25	15	3.80	0.33				
7	25	35	7.30	0.79				
8	15	25	4.90	0.35				
9	35	25	5.90	0.71	0.0134	0.3088	0.4040	0.2738
10	54	28	9.66	0.61	0.0004	0.3044	0.4962	0.1990
11	17	45	6.30	0.52	0.0000	0.2485	0.6163	0.1351
12	8	55	8.89	0.34	0.0000	0.2293	0.6790	0.0917

Table 6.3.2
Weights used in the criterion function ϕ at different stages;discrimination among nonlinear bivariate models

Sequential stage k	Model v	Model u 1	2	3
1	2	0.017669		
	3	0.013503	0.311402	
	4	0.019924	0.361349	0.276152
2	2	0.000719		
	3	0.000441	0.367036	
	4	0.001101	0.390869	0.239833
3	2	0.000025		
	3	0.000010	0.345698	
	4	0.000046	0.466225	0.187995
4	2	0.000001		
	3	0.000000	0.387070	
	4	0.000002	0.458196	0.154731

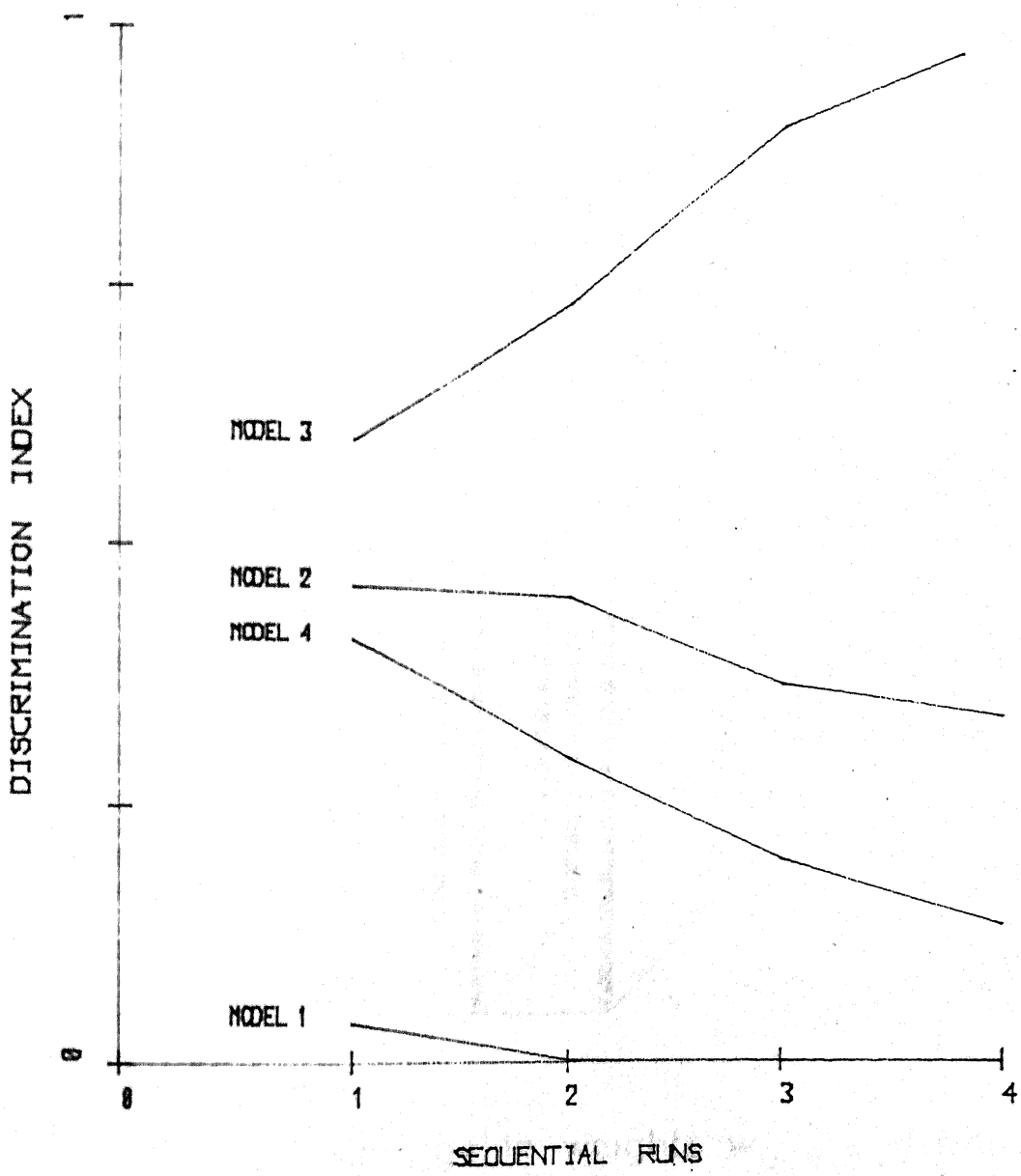

FIGURE 6.3.1 Performance of the proposed procedure in discriminating among nonlinear bivariate models. True model : Model 1.

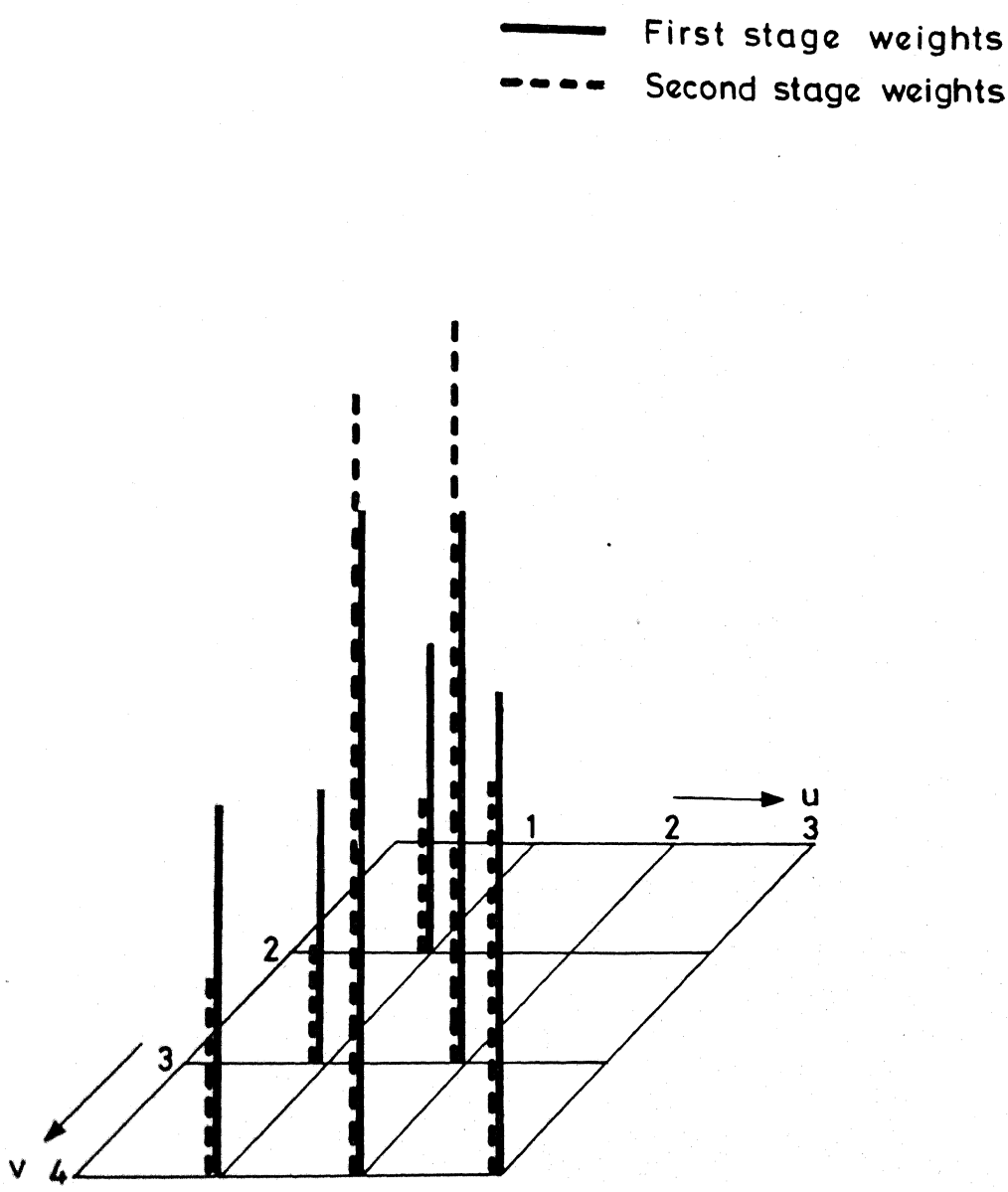

FIGURE 6.3.2 Change in weights $w_{u,v;k}$, from first to second stage; nonlinear bivariate models.

Table 6.3.3

Sequential discrimination among nonlinear bivariate models : the procedure of Buzzi et al.

(True model : $M^{(1)}$)

Run k	Input variables ξ_{1k}	ξ_{2k}	Responses y_{1k}	y_{2k}	Chi - Square statistics X_1^2	X_2^2	X_3^2	X_4^2
1-9	(s e e Table 6.3.1)				15.1	14.7	24.3	14.4
10	55.0	32.8	9.15	0.93	15.9	26.5	58.3	15.6
11	55.0	55.0	13.74	1.34	18.9	72.2	68.0	23.8
12	10.6	55.0	6.00	0.70	21.8	81.0	101.0	29.4
13	16.0	55.0	8.20	0.84	21.8	81.5	151.0	34.8
14	5.0	5.0	0.45	0.04				
15	5.0	55.0	3.27	0.44	Points generated for precise estimation			
16	55.0	5.0	1.47	1.13				
17	55.0	55.0	14.00	1.37	25.1	114.6	117.3	36.4
18	10.6	55.0	5.50	0.55	33.0	108.0	185.0	52.8
19	16.0	55.0	8.40	0.70	35.7	119.0	259.0	67.3

$$\chi^2_u = \sum_{k=1}^{n} (\underset{\sim}{y}_k - \underset{\sim}{y}_k^{(u)})' \Sigma^{-1} (\underset{\sim}{y}_k - \underset{\sim}{y}_k^{(u)})$$

(with known Σ), used by them for this purpose, indicate that all the models are adequate. The addition of the responses (9.15, 0.93) corresponding to the setting (55.0, 32.8) of $\underset{\sim}{\xi}_{10}$ renders $M^{(3)}$ as an inadequate model with $\chi^2_u = 58.3$. According to their scheme this model is dropped for the purpose of designing the next experiment. In fact, $M^{(3)}$ continues to be inadequate throughout the sequential procedure, as can be seen through Table 6.3.3. Similarly, $M^{(2)}$, too, once dropped after 11 experiments never got a chance to be included in the design criterion. This way, only $M^{(1)}$ and $M^{(4)}$ participated in designing new experiments. These models continued to be close rivals till the 19th experiment, when this procedure is stopped according to the prescribed condition. The model $M^{(1)}$ is declared as the correct model.

It can be clearly seen that this decision which could be possible after 19 experiments through the procedure of Buzzi et al. has been made after 12 experiments only, through the procedure proposed in this work. In addition to the faster convergence, the present method has also an advantage of handling situations when Σ is unknown, as has been assumed in the present case.

6.3.2 Example : Nonlinear Models; Unknown; Unequal Covariance Matrices of Errors

In this example we consider an hypothetical biresponse system for which the model:

$$M^{(0)}: \quad \eta_1^{(0)} = \theta_1^{(0)}\theta_3^{(0)}\xi_1\xi_2/(1+\theta_3^{(0)}\xi_1)$$

$$\eta_2^{(0)} = \theta_2^{(0)}\theta_3^{(0)}\xi_1\xi_2/(1+\theta_3^{(0)}\xi_1)^2 \qquad (6.3.2)$$

with $\theta_1^{(0)} = 0.0005$, $\theta_2^{(0)} = 0.16$, and $\theta_3^{(0)} = 15$, is the true model. So far as simulation of responses from this system is concerned we use the equations

$$y_1 = \eta_1^{(0)} + \varepsilon_1$$
$$y_2 = \eta_2^{(0)} + \varepsilon_2 ,$$

where the independent random errors ε_1, ε_2 are distributed normally with means, 0.0, 0.0, and standard deviations, 3.162×10^{-6} and 1.0×10^{-4}, respectively. This set-up for generation of data is the one that has been used by Hunter and Wichern (1966). The data generated through $M^{(0)}$ is planned to be used for discriminating among three bivariate models:

$$M^{(1)}: \quad \eta_1^{(1)} = \theta_1^{(1)}\theta_3^{(1)}\theta_4^{(1)}\xi_1\xi_2/(1+\theta_3^{(1)}\xi_1+\theta_4^{(1)}\xi_2)^2$$

$$\eta_2^{(1)} = \theta_2^{(1)}\theta_3^{(1)}\theta_4^{(1)}\xi_1\xi_2/(1+\theta_3^{(1)}\xi_1+\theta_4^{(1)}\xi_2)^2 ,$$

$$M^{(2)}: \quad \eta_1^{(2)} = \theta_1^{(2)}\theta_3^{(2)}\theta_4^{(2)}\xi_1\xi_2/(1+\theta_3^{(2)}\xi_1+\theta_4^{(2)}\xi_2)^2$$

$$\eta_2^{(2)} = \theta_2^{(2)}\theta_3^{(2)}\theta_4^{(2)}\xi_1\xi_2/(1+\theta_3^{(2)}\xi_1+\theta_4^{(2)}\xi_2) ,$$

Table 6.3.4
Discrimination index and Posterior probability in discriminating among nonlinear bivariate models
(True model : $M^{(3)}$)

Run k	Input Variables ξ_{1k}	ξ_{2k}	Response y_{1k}	y_{2k}	Discrimination criteria $D_k^{(1)}$ ($P_k^{(1)}$)	$D_k^{(2)}$ ($P_k^{(2)}$)	$D_k^{(3)}$ ($P_k^{(3)}$)
1	1.0	2.0	0.00094	0.01861			
2	4.0	2.0	0.00098	0.00520			
3	1.0	3.0	0.00140	0.02820			
4	4.0	3.0	0.00148	0.00780	0.5578 (0.0390)	0.4415 (0.0000)	0.0007 (0.9610)
5	5.0	3.0	0.01480	0.00620	0.5786 (0.0010)	0.4214 (0.0000)	0.0000 (0.9990)

$$M^{(3)}: \quad \eta_1^{(3)} = \theta_1^{(3)}\theta_3^{(3)}\theta_4^{(4)}\xi_1\xi_2 / (1 + \theta_3^{(3)}\xi_1)$$

$$\eta_2^{(3)} = \theta_2^{(3)}\theta_3^{(3)}\theta_4^{(3)}\xi_1\xi_2 / (1 + \theta_3^{(3)}\xi_1)^2 .$$

Initially, four experiments are selected according to a 2^2 factorial design and $M^{(0)}$ is used to simulate observations on the responses y_1, y_2, as shown in Table 6.3.4. These values when utilized in (2.5.27) and subsequently in (2.4.1) yield 0.5578, 0.4415, and 0.0067 as the values of the discrimination index for $M^{(1)}$, $M^{(2)}$, and $M^{(3)}$, respectively. These values show enough evidence in favour of $M^{(3)}$. However, we design some more points. To that purpose the independent variables ξ_1, ξ_2 are constrained within their operability region, $\{0 \leq \xi_1, \xi_2 \leq 5\}$. Before we proceed to design the 5th experiment we must decide on the weights. In this connection we notice that the weights calculated through the strategy proposed here are appropriate, in that, the pair consisting of models $M^{(1)}$ and $M^{(2)}$, which have been found to be very close to each other, is given the highest weight of 0.996562. Since other pairs, i.e., $(M^{(1)}, M^{(3)})$ and $(M^{(2)}, M^{(3)})$ consist of almost equidistant models, there is not much difference in their respective weights; being, 0.001519 and 0.001919.

Since the covariance matrices in the present case have been assumed to be unknown, we use h_3 of (4.4.16) in the criterion function $\phi(\xi_{4+1})$. The maximization of the resulting function gives (5,3) as the next setting of the experiment which in turn produces 0.00148 and 0.00622 as the values of the two

responses, through $M^{(0)}$. The values of the discrimination index which results from the sample of (4+1) observations shows that while the gap between $M^{(1)}$ and $M^{(2)}$ is widened, model $M^{(3)}$, too, is further pulled apart from the other residual models, as can be appreciated through the values of $D_5^{(u)}$, u = 1,2,3, given in Table 6.3.4. Besides, the low value, 0.12289×10^{-5} of $D_5^{(3)}$ establishes the superiority of $M^{(3)}$. At this stage a look at the weights for the 6th experiment, if required, clearly suggests that $M^{(1)}$ need not be further discriminated from $M^{(2)}$ and $M^{(3)}$, since the pairs $(M^{(1)},M^{(2)})$ and $(M^{(1)},M^{(3)})$ require as low weights as 0.000003 and 0.000004, respectively. This fact, together with the low order of $D_5^{(3)}$ suggests the termination of the sequential procedure. Thus with the help of a single additional experiment $M^{(3)}$ is fairly well identified as being the best model.

Hunter and Wichern (1966) have used the Box-Hill (1967) procedure in simulating discrimination among the three models considered in this example. Based on the same preliminary design and using the identical hypothetical system, the posterior probabilities of the three models at two sequential stages are listed in Table 6.3.4 as bracketed values. In fact, after 6 experiments no further discrimination is required as the probability of one of the models; namely, $M^{(3)}$ has risen to 0.999, which carries enough evidence in favour of this model. In this example the convergence through the Box-Hill procedure, too, happens to be quite fast. However, through our procedure we could still save one run of the experiment.

A P P E N D I X A

COMBINING TWO QUADRATIC FORMS

Lemma A.1. If Q_1 and Q_2 are two quadratic forms in $\underset{\sim}{x}$; $Q_i = (\underset{\sim}{x}-\underset{\sim}{v}_i)' M_i^{-1}(\underset{\sim}{x}-\underset{\sim}{v}_i)$, $i = 1,2$, $\underset{\sim}{x},\underset{\sim}{v}_i$ as $r \times 1$ vectors and M_i^{-1} as the inverse of $r \times r$ symmetric matrix M_i, $i = 1,2$, then

$$Q_1 + Q_2 =$$

$$(\underset{\sim}{x}-\underset{\sim}{w})' M^*(\underset{\sim}{x}-\underset{\sim}{w}) + (\underset{\sim}{v}_1-\underset{\sim}{v}_2)' (M_1+M_2)^{-1} (\underset{\sim}{v}_1-\underset{\sim}{v}_2), \qquad (A.1.1)$$

where

$$\underset{\sim}{w} = (M_1+M_2)^{-1} (M_2\underset{\sim}{v}_1 + M_1\underset{\sim}{v}_2), \qquad (A.1.2)$$

$$M^* = M_1^{-1}(M_1+M_2) M_2^{-1} \qquad (A.1.3)$$

Proof:

Let $M_i^{-1} = A_i$, $i = 1,2$, and $A = A_1+A_2$

Let $(A_1+A_2)^{-1}$ $(=A^{-1})$ denote the inverse of A_1+A_2 $(= A)$

Then Q_1+Q_2 can be written as

$$Q_1+Q_2 = (\underset{\sim}{x}-\underset{\sim}{v}_1)'A_1 (\underset{\sim}{x}-\underset{\sim}{v}_1) + (\underset{\sim}{x}-\underset{\sim}{v}_2)' A_2(\underset{\sim}{x}-\underset{\sim}{v}_2) \qquad (A.1.4)$$

Simplifying and rearranging the terms on the right hand side of (A.1.4), we get

$$\begin{aligned} Q_1+Q_2 &= \underset{\sim}{x}'A \underset{\sim}{x} - 2\underset{\sim}{x}' (A_1\underset{\sim}{v}_1 + A_2\underset{\sim}{v}_2) + \underset{\sim}{v}_1' A_1\underset{\sim}{v}_1 + \underset{\sim}{v}_2' A_2 \underset{\sim}{v}_2 \\ &= \underset{\sim}{x}' A \underset{\sim}{x}'- 2\underset{\sim}{x}' A \underset{\sim}{w} + \underset{\sim}{w}' A \underset{\sim}{w} + \underset{\sim}{v}_1'A_1\underset{\sim}{v}_1 + \underset{\sim}{v}_2' A_2 \underset{\sim}{v}_2 - \underset{\sim}{w}' A \underset{\sim}{w} \\ &= (\underset{\sim}{x}-\underset{\sim}{w})' A(\underset{\sim}{x}-\underset{\sim}{w}) + \underset{\sim}{v}_1' A_1 \underset{\sim}{v}_1 + \underset{\sim}{v}_2' A_2 \underset{\sim}{v}_2 - \underset{\sim}{w}' A \underset{\sim}{w}, \end{aligned}$$

where $\underset{\sim}{w} = A^{-1}(A_1\underset{\sim}{v}_1 + A_2\underset{\sim}{v}_2)$

But

$$\underset{\sim}{w}' A \underset{\sim}{w} = (A_1\underset{\sim}{v}_1 + A_2\underset{\sim}{v}_2)' A^{-1} A A^{-1}(A_1\underset{\sim}{v}_1 + A_2\underset{\sim}{v}_2)$$

$$= [A_1(\underset{\sim}{v}_1-\underset{\sim}{v}_2) + A\underset{\sim}{v}_2]' A^{-1}[A\underset{\sim}{v}_1 - A_2(\underset{\sim}{v}_1-\underset{\sim}{v}_2)]$$

$$= (\underset{\sim}{v}_1-\underset{\sim}{v}_2)' A_1A^{-1}A\underset{\sim}{v}_1 - (\underset{\sim}{v}_1-\underset{\sim}{v}_2)' A_1A^{-1}A_2(\underset{\sim}{v}_1-\underset{\sim}{v}_2) +$$

$$\underset{\sim}{v}_2' A A^{-1} A \underset{\sim}{v}_1 - \underset{\sim}{v}_2 A A^{-1} A_2(\underset{\sim}{v}_1-\underset{\sim}{v}_2)$$

$$= (\underset{\sim}{v}_1-\underset{\sim}{v}_2)' A_1\underset{\sim}{v}_1 - (\underset{\sim}{v}_1-\underset{\sim}{v}_2)'A_1A^{-1}A_2(\underset{\sim}{v}_1-\underset{\sim}{v}_2)+\underset{\sim}{v}_2'A\underset{\sim}{v}_1-\underset{\sim}{v}_2'A_2(\underset{\sim}{v}_1-\underset{\sim}{v}_2)$$

Thus

$$\underset{\sim}{w}' A \underset{\sim}{w} = \underset{\sim}{v}_1' A_1 \underset{\sim}{v}_1 + \underset{\sim}{v}_2' A_2 \underset{\sim}{v}_2 - (\underset{\sim}{v}_1-\underset{\sim}{v}_2)' A_1 A^{-1} A_2(\underset{\sim}{v}_1-\underset{\sim}{v}_2) \qquad (A.1.6)$$

Substituting (A.1.6) in (A.1.5) we have

$$Q_1+Q_2 = (\underset{\sim}{x}-\underset{\sim}{w})'A(\underset{\sim}{x}-\underset{\sim}{w}) + (\underset{\sim}{v}_1-\underset{\sim}{v}_2)' A_1A^{-1}A_2(\underset{\sim}{v}_1-\underset{\sim}{v}_2) \qquad (A.1.7)$$

When we reconsider the substitution $A_i = M_i^{-1}$, $i = 1,2$

we also use

$$A = (M_1^{-1}+M_2^{-1}) \text{ and } A^{-1} = (M_1^{-1}+M_2^{-1})^{-1}$$

Besides, the symmetry of M_1, M_2 and consequently of M_1^{-1} and M_2^{-1} enable us to use the following relations

$$M_1(M_1+M_2)^{-1} M_2 = (M_1^{-1}+M_2^{-1})^{-1}$$

and $\quad (M_1^{-1}+M_2^{-1}) = M_1^{-1}(M_1+M_2)M_2^{-1}.$

As a result we have

$$A = M_1^{-1}(M_1+M_2) M_2^{-1} = M^* \text{ (say)} \qquad (A.1.8)$$

$$A^{-1} = M_1(M_1+M_2)^{-1} M_2$$

and from (A.1.7) we get

$$Q_1+Q_2 = (\underset{\sim}{x}-\underset{\sim}{w})' M^*(\underset{\sim}{x}-\underset{\sim}{w}) + (\underset{\sim}{v}_1-\underset{\sim}{v}_2)' (M_1+M_2)^{-1}(\underset{\sim}{v}_1-\underset{\sim}{v}_2), \quad (A.1.9)$$

where

$$\begin{aligned}\underset{\sim}{w} &= (A_1+A_2)^{-1} (A_2\underset{\sim}{v}_1 + A_2\underset{\sim}{v}_2) \\ &= (M_1^{-1}+M_2^{-1})^{-1} (M_1^{-1}\underset{\sim}{v}_1 + M_2^{-1}\underset{\sim}{v}_2) \\ &= M_1(M_1+M_2)^{-1} M_2(M_1^{-1}\underset{\sim}{v}_1 + M_2^{-1}\underset{\sim}{v}_2),\end{aligned}$$

i.e.,

$$\underset{\sim}{w} = (M_1+M_2)^{-1} (M_2\underset{\sim}{v}_1 + M_1\underset{\sim}{v}_2). \quad (A.1.10)$$

This proves the Lemma.

<u>Corollary A.1</u> We now replace the vectors $\underset{\sim}{x},\underset{\sim}{v}_1,\underset{\sim}{v}_2$ and the matrices M_1,M_2 by scalars x,v_1,v_2 and m_1,m_2, respectively, so that M_i^{-1} will be replaced by $1/m_i$ $(m_i \neq 0)$ and $Q_i = (x-v_i)^2/m_i$, $i = 1,2$. As a result we have

$$Q_1+Q_2 = \frac{(x-w)^2}{m^*} + \frac{(v_1-v_2)^2}{m_1+m_2}, \quad (A.1.11)$$

where $m^* = \frac{m_1 m_2}{(m_1+m_2)}$ and $w = \frac{m_2v_1 + m_1v_2}{m_1+m_2}$. (A.1.12)

A P P E N D I X B

SOME RESULTS OF MATRIX ALGEBRA

Lemma B.1 Let A_1 be an $r \times r$ positive definite symmetric matrix and A_2 be an $r \times r$ non-negative definite symmetric matrix. Then for sufficiently small α

$$\log |I - \alpha A_1 A_2| = -\sum_{b=1}^{\infty} \frac{\alpha^b}{b} \operatorname{tr}[(A_1 A_2)^b] \cdot \qquad (B.1.1)$$

Proof :

Let $A_1 A_2 = M$.

Case 1 : Suppose $A_1 A_2 = A_2 A_1$.

Then A_1 and A_2 being symmetric, M is also symmetric and we can write

$$(I-\alpha M) = T[\operatorname{diag}\{1-\alpha\gamma_1)(1-\alpha\gamma_2),\ldots,(1-\alpha\gamma_r)\}]T' ,$$

where $\gamma_1, \gamma_2, \ldots, \gamma_r$ are the eigen values of M, T is some suitable transformation such that $T'T = I$, and I is an $r \times r$ identity matrix.

Therefore

$$|I-\alpha M| = \prod_{i=1}^{r} (1-\alpha\gamma_i)$$

and

$$\log|I-\alpha M| = \sum_{i=1}^{r} \log (1-\alpha\gamma_i) \cdot$$

Since α is supposed to be sufficiently small, we can expand $\log (1-\alpha\gamma_i)$ and write

$$\log |I-\alpha M| = -\sum_{i=1}^{r} \left[\sum_{b=1}^{\infty} \frac{\alpha^b}{b} \gamma_i^b\right]$$

$$= -\sum_{b=1}^{\infty} \frac{\alpha^b}{b} \operatorname{tr}(M^b),$$

i.e., $$\log |I-\alpha A_1 A_2| = -\sum_{b=1}^{\infty} \frac{\alpha^b}{b} \operatorname{tr}\left[(A_1 A_2)^b\right].$$

Case ii If $M(= A_1 A_2)$ has distinct eigen values, then we can write

$$(I-\alpha M) = T[\operatorname{diag}\{(1-\alpha\gamma_1)(1-\alpha\gamma_2),..,(1-\alpha\gamma_r)\}]\, T^{-1},$$

where the matrix T can be formed by taking the eigen vectors of $(I-\alpha M)$ as the columns of T.

Case iii If the eigen values of M are not distinct then using the Jordan Canonical form, we have [Bellman (1977)]

$$I-\alpha M = T[\operatorname{diag}\{L_{m_1}(\gamma_1), L_{m_2}(\gamma_2),\ldots,L_{m_s}(\gamma_s)\}]T^{-1},$$

where γ_j is an eigen value of multiplicity m_j, $j = 1,2,\ldots,s$, with $r = \sum_{j=1}^{s} m_j$ and $L_{m_j}(\gamma_j)$ is an $m_j \times m_j$ matrix, given by

$$L_{m_j}(\gamma_j) = \begin{bmatrix} (1-\alpha\gamma_j) & 1 & 0 & \ldots & 0 & 0 \\ 0 & (1-\alpha\gamma_j) & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ \cdot & \cdot & \cdot & \ldots & (1-\alpha\gamma_j) & 1 \\ 0 & 0 & 0 & \ldots & 0 & (1-\alpha\gamma_j) \end{bmatrix}.$$

Therefore,

$$|I-\alpha M| = \prod_{j=1}^{s} (1-\alpha\gamma_j)^{m_j}.$$

Taking logarithm on both sides, we get

$$\log |I - \alpha M| = \sum_{j=1}^{s} m_j \log (1 - \alpha \gamma_j) .$$

Since α is sufficiently small we can expand $\log (1 - \alpha \gamma_j)$ and obtain

$$\log |I - \alpha M| = - \sum_{j=1}^{s} m_j \sum_{b=1}^{\infty} \frac{\alpha^b}{b} \gamma_j^b$$

$$= - \sum_{b=1}^{\infty} \frac{\alpha^b}{b} \sum_{j=1}^{s} m_j \gamma_j^b ,$$

i.e.,

$$\log |I - \alpha A_1 A_2| = - \sum_{b=1}^{\infty} \frac{\alpha^b}{b} \operatorname{tr} [(A_1 A_2)^b] .$$

Corollary B.1 If a_1 and a_2 are two numbers such that $a_1 > 0$ $a_2 \geq 0$, then from (B.1.1) we get

$$\log (1-\alpha a_1 a_2) = - \sum_{b=1}^{\infty} \frac{\alpha^b}{b} (a_1 a_2)^b, \qquad (B.1.2)$$

for sufficiently small $\alpha > 0$.

Lemma B.2 If B_1 and B_2 are two matrices such that $B_1 B_2 = B_2 B_1$, then

$$(B_1+B_2)^a = \sum_{b=0}^{\infty} \binom{a}{b} B_1^b B_2^{a-b}. \qquad (B.2.1)$$

Proof : By induction.

The result is obvious for $a = 1$.

When $a = 2$,

$$(B_1+B_2)^2 = B_1^2 + 2B_1B_2 + B_2^2$$

$$= \sum_{b=0}^{2} \binom{2}{b} B_1^b B_2^{2-b}.$$

Let the identity (B.2.1) be true for a = k. Then

$$(B_1+B_2)^k = \sum_{b=0}^{k} \binom{k}{b} B_1^k B_2^{k-b}$$

$$= B_1^k+\binom{k}{k-1}B_1^{k-1}B_2+\ldots+\binom{k}{j-1}B_1^{j-1}B_2^{k-j+1}+\binom{k}{j}B_1^jB_2^{k-j} + \ldots + B_2^k.$$

Multiplying both sides by (B_1+B_2), we get

$$(B_1+B_2)^{k+1} =$$

$$B_1^{k+1}+\binom{k}{k-1}B_1^kB_2+\ldots+\binom{k}{j-1}B_1^jB_2^{k-j+1}+\binom{k}{j}B_1^{j+1}B_2^{k-j} +\ldots+ B_1B_2^k$$

$$+ B_2B_1^k+\binom{k}{k-1}B_1^kB_2^2 +\ldots+ \binom{k}{j-1} B_1^{j-1}B_2^{k-j+2}+\binom{k}{j}B_1^jB_2^{k-j+1}+\ldots+B_2^{k+1}.$$

Collecting terms of like powers, we get

$$(B_1+B_2)^{k+1} =$$

$$B_1^{k+1}+\left[\binom{k}{k-1}+\binom{k}{k}\right] B_1^kB_2+\ldots+\left[\binom{k}{j-1}+\binom{k}{j}\right] B_1^jB_2^{k-j+1} +\ldots$$

$$+ \left[\binom{k}{k} + \binom{k}{k-1}\right] B_1B_2^k + B_2^{k+1}.$$

This shows that the coefficient of the general term (i.e. the jth term) in this expansion is

$$\binom{k}{j-1} + \binom{k}{j} = \binom{k+1}{j}.$$

Thus, we can write

$$(B_1+B_2)^{k+1} = \sum_{j=0}^{k+1} \binom{k+1}{j} B_1^j B_2^{k+1-j},$$

showing that the result is true for a = k+1.

Hence, by mathematical induction

$$(B_1+B_2)^a = \sum_{b=0}^{a} \binom{a}{b} B_1^b B_2^{a-b}.$$

Corollary B.2 If the matrices B_1 and B_2 are replaced by the scalars b_1 and b_2, respectively, we have for the positive integer a

$$(b_1+b_2)^a = \sum_{i=0}^{a} \binom{a}{i} b_1^i b_2^{a-i}. \qquad (B.2.2)$$

Lemma B.3 If B_1 and B_2 are two matrices such that $B_1B_2 = B_2B_1$, then for sufficiently small α_1, α_2

$$(I-\alpha_1B_1-\alpha_2B_2)^{-1} = \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} \binom{a+b}{a} \alpha_1^a \alpha_2^b B_1^a B_2^b. \qquad (B.3.1)$$

Proof : Let $\alpha_1B_1 + \alpha_2B_2 = A$.

Then $$(I-A)^{-1} = I + A + A^2 + \ldots + A^j + \ldots .$$

$$= \sum_{j=0}^{\infty} A^j.$$

But, $A^j = (\alpha_1B_1+\alpha_2B_2)^j$ can be expanded using Lemma B.1 so that

$$A^j = \sum_{k=0}^{j} \binom{j}{k} \alpha_1^k \alpha_2^{j-k} B_1^k B_2^{j-k}.$$

Consequently, we can write $(I-A)^{-1}$ as

$$(I-A)^{-1} = \sum_{j=0}^{\infty} \sum_{k=0}^{j} \binom{j}{k} \alpha_1^k \alpha_2^{j-k} B_1^k B_2^{j-k}$$

and if we let k = a and j-k = b, then

$$(I-\alpha_1B_1-\alpha_2B_2)^{-1} = \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} \binom{a+b}{a} \alpha_1^a \alpha_2^b B_1^a B_2^b.$$

This proves the identity (B.3.1).

Corollary B.3 If b_1 and b_2 are any two scalars, then for sufficiently small α_1, α_2

$$(1-\alpha_1 b_1-\alpha_2 b_2)^{-1} = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \binom{i+j}{i} \alpha_1^i \alpha_2^j b_1^i b_2^j \qquad (B.3.2)$$

as can be obtained from (B.3.1) by replacing the matrices I, B_1, and B_2 by 1, b_1, and b_2, respectively.

APPENDIX C

JOINT CUMULANT GENERATING FUNCTION

C.1 Univariate Case

Lemma C.1. Let Y be a normal random variable with mean μ and variance λ^{-1}. Then the joint cumulant generating function for Q_1 and Q_2; $Q_i = \lambda_i(y-\mu_i)^2$, $i = 1,2$, is given by

$$k(w_1,w_2) = \sum_{a=1}^{\infty} [2^{a-1}(a-1)!\ (1+a\alpha d_1^2)\ (\frac{\lambda_1}{\lambda})^a]\ \frac{w_1^a}{a!}$$

$$+ \sum_{b=1}^{\infty} [2^{b-1}(b-1)!\ (1+b\alpha d_2^2)\ (\frac{\lambda_2}{\lambda})^b]\ \frac{w_2^b}{b!}$$

$$+ \sum_{a=1}^{\infty}\sum_{b=1}^{\infty} [2^{a+b-1}(a+b-2)!\ \{(a+b-1)+a(a-1)\ \lambda d_1^2 +$$

$$b(b-1)\lambda\ d_2^2 + 2ab\lambda\ d_1 d_2\}\ (\frac{\lambda_1}{\lambda})^a(\frac{\lambda_2}{\lambda})^b]\ \frac{w_1^a}{a!}\ \frac{w_2^b}{b!}\ , \qquad (C.1.1)$$

for some real w_1,w_2, where $d_i = \mu - \mu_i$, $i = 1,2$; $\lambda = \lambda_1+\lambda_2$.

Proof : Since Y is distributed as $N(\mu,\lambda^{-1})$ the joint cumulant generating function for Q_1 and Q_2, by definition, is given by

$$k(w_1,w_2) = \log\ [(\lambda/2\pi)^{1/2} \int_{R^1} \exp(w_1Q_1+w_2Q_2)\exp(-\tfrac{1}{2}\ Q)dy], \qquad (C.1.2)$$

where $Q = \lambda(y-\mu)^2$.

Now,

$$w_1Q_1+w_2Q_2-\tfrac{1}{2}Q = \tfrac{\lambda}{2}\{2c_1(y-\mu_1)^2+2c_2(y-\mu_2)^2-(y-\mu)^2\},$$

where $c_i = \dfrac{w_i\lambda_i}{\lambda}$, $i = 1,2.$

This can also be written as

$$w_1Q_1+w_2Q_2-\tfrac{1}{2}Q = -\tfrac{1}{2}(\lambda\delta z^2) + 2\lambda(c_1d_1+c_2d_2)z + (c_1d_1^2+c_2d_2^2), \qquad (C.1.3)$$

where $\delta = (1-2c_1-2c_2)$, $z = (y-\mu)$, and $d_i = (\mu-\mu_i)$, $i=1,2.$

Substituting (C.1.3) in (C.1.2) we get

$$k(w_1,w_2) = -\tfrac{1}{2}\log\delta + \log\left[(\lambda\delta/2\pi)^{1/2} \int_{R^1} \exp\{2\lambda(c_1d_1+c_2d_2)z\} \exp(-\tfrac{1}{2}\lambda\delta z^2)dz\right] + \lambda(c_1d_1^2+c_2d_2^2),$$

i.e.,

$$k(w_1,w_2) = -\tfrac{1}{2}\log\delta + 2\lambda(c_1d_1+c_2d_2)^2\delta^{-1} + \lambda(c_1d_1^2+c_2d_2^2). \qquad (C.1.4)$$

Now,

$$c_1d_1^2+c_2d_2^2 = (c_1d_1^2+c_2d_2^2)(1-2c_1-2c_2)\delta^{-1}$$

$$= \{c_1d_1^2+c_2d_2^2-c_1^2d_1^2-c_2^2d_2^2 - 2c_1c_2(d_1^2+d_2^2)\}\delta^{-1}.$$

So that from (C.1.3) we obtain

$$k(w_1,w_2) = -\tfrac{1}{2}\log\delta + \lambda c_1d_1^2\delta^{-1}+\lambda c_2d_2^2\delta^{-1}-2\lambda c_1c_2(d_1-d_2)^2\delta^{-1}. \qquad (C.1.5)$$

Using equations (B.1.2) and (B.2.2) from Lemma B.1, we can express the first component of (C.1.4) as

$$-\frac{1}{2}\log\delta = \sum_{a=1}^{\infty}\frac{2^{a-1}}{a}\sum_{i=0}^{a}\binom{a}{i}c_1^i c_2^{a-i}$$

$$= \sum_{a=1}^{\infty}2^{a-1}\left[\frac{c_2^a}{a} + \sum_{b=1}^{\infty}\frac{2^b}{a+b}\binom{a+b}{a}c_1^a c_2^b + \frac{c_1^a}{a}\right]$$

$$= \sum_{a=1}^{\infty}\frac{2^{a-1}}{a}c_1^a + \sum_{b=1}^{\infty}\frac{2^{b-1}}{b}c_2^b + \sum_{a=1}^{\infty}\sum_{b=1}^{\infty}2^{a+b-1}\frac{(a+b-1)!}{a!\,b!}c_1^a c_2^b. \tag{C.1.6}$$

Also, we have, from relation (B.3.2) of Lemma B.3,

$$\delta^{-1} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}2^{i+j}\binom{i+j}{i}c_1^i c_2^j$$

through which we can express the other components of (C.1.5) in the form of power series, e.g.,

$$\lambda c_1 d_1^2\delta^{-1} = \lambda d_1^2\left[\sum_{i=0}^{\infty}2^i c_1^{i+1}\sum_{j=0}^{\infty}2^j\binom{i+j}{i}c_2^j\right]$$

$$= \lambda d_1^2\left[\sum_{a=1}^{\infty}2^{a-1}c_1^a + \sum_{a=1}^{\infty}\sum_{b=1}^{\infty}2^{a+b-1}\frac{(a+b-1)!}{(a-1)!\,b!}c_1^a c_2^b\right]. \tag{C.1.7}$$

Similarly,

$$\lambda c_2 d_2^2\delta^{-1} = d_2^2\left[\sum_{b=1}^{\infty}2^{b-1}c_2^b + \sum_{a=1}^{\infty}\sum_{b=1}^{\infty}2^{a+b-1}\frac{(a+b-1)!}{a!(b-1)!}c_1^a c_2^b\right]. \tag{C.1.8}$$

Besides,

$$\lambda c_1 c_2 (d_1-d_2)^2 \delta^{-1} = \lambda(d_1^2+d_2^2 - 2d_1d_2) \times$$

$$\left[\sum_{a=1}^{\infty} \sum_{b=1}^{\infty} 2^{a+b-2} \frac{(a+b-2)!}{(a-1)!(b-1)!} c_1^a c_2^b\right]. \qquad \text{(C.1.9)}$$

Substituting (C.1.6) through (C.1.9) in (C. 1.5) and rearranging the terms appropriately, we get

$$k(w_1,w_2) = \sum_{a=1}^{\infty} \frac{2^{a-1}}{a} (1+a\,\lambda d_1^2)\, c_1^a + \sum_{b=1}^{\infty} \frac{2^{b-1}}{b} (1+b\,\lambda d_2^2) c_2^b$$

$$+ \sum_{a=1}^{\infty} \sum_{b=1}^{\infty} 2^{a+b-1} \frac{(a+b-2)!}{a!\,b!}$$

$$\{(a+b-1) + a(a+b-1) + b(a+b-1) - ab\}\, c_1^a c_2^b.$$

Simplifying further and substituting back $c_i = (w_i\lambda_i)/\lambda$, $i=1,2$, we finally have

$$k(w_1,w_2) = \sum_{a=1}^{\infty} \left[2^{a-1}\,(a-1)!\,(1+a\,\lambda d_1^2)\,\left(\frac{\lambda_1}{\lambda}\right)^a\right] \frac{w_1^a}{a!}$$

$$+ \sum_{b=1}^{\infty} \left[2^{b-1}(b-1)!\,(1+b\,\lambda d_2^2)\,\left(\frac{\lambda_2}{\lambda}\right)^b\right] \frac{w^b}{b!}$$

$$+ \sum_{a=1}^{\infty} \sum_{b=1}^{\infty} \left[2^{a+b-1}(a+b-2)!\,\{(a+b-1)+a(a-1)\lambda d_1^2+b(b-1)\lambda d_2^2\right.$$

$$\left. + 2ab\,\lambda d_1 d_2\}\,\left(\frac{\lambda_1}{\lambda}\right)^a \left(\frac{\lambda_2}{\lambda}\right)^b\right] \frac{w_1^a}{a!}\,\frac{w_2^b}{b!}. \qquad \text{(C.1.10)}$$

The function $k(w_1,w_2)$ given by (C.1.10) can be used to evaluate any required cumulant : the cumulant k_{ao} is the coefficient of $(w_1^a/a!)$, the cumulant k_{ob} is the coefficient of $(w_2^b/b!)$ and

the cumulant k_{ab} is the coefficient of $(w_1^a w_2^b/a!b!)$ in the above expression. Thus

$$k_{ao} = 2^{a-1}(a-1)!\ (1+a\,\lambda d_1^2)\ (\frac{\lambda_1}{\lambda})^a\ , \qquad (C.1.11)$$

$$k_{ob} = 2^{b-1}(b-1)!\ (1+b\,\lambda d_2^2)\ (\frac{\lambda_2}{\lambda})^b\ , \qquad (C.1.12)$$

$$k_{ab} = 2^{a+b-1}(a+b-2)!\ \{(a+b-1) + a(a-1)\,\lambda d_1^2 + b(b-1)\,\lambda d_2^2 + 2ab\,\lambda d_1 d_2\}\ (\frac{\lambda_1}{\lambda})^a\ (\frac{\lambda_2}{\lambda})^b\ , \qquad (C.1.13)$$

where $d_i = (\mu - \mu_i)$, $i = 1,2$.

C.2 Multivariate Case

Lemma C.2. Let $\underset{\sim}{Y}$ be a random vector having an r-variate normal distribution : $N_r(\underset{\sim}{\mu}, \Lambda^{-1})$ and let Q_1, Q_2 be two quadratic forms in $\underset{\sim}{Y}$ such that $Q_i = (\underset{\sim}{Y} - \underset{\sim}{\mu}_i)'\ \Lambda_i\ (\underset{\sim}{Y} - \underset{\sim}{\mu}_i)$, $i = 1,2$, with Λ_1, Λ_2 as symmetric matrices and $\Lambda = \Lambda_1 + \Lambda_2$. Then the joint cumulant generating function of Q_1 and Q_2 is

$$\begin{aligned} k(w_1,w_2) = {} & \sum_{a=1}^{\infty} 2^{a-1}(a-1)!\ \{tr(\Sigma_1^a) + a\underset{\sim}{d}_1'\ \Lambda^{-1}\ \Sigma_1^a \underset{\sim}{d}_1\}\ \frac{w_1^a}{a!} \\ & + \sum_{b=1}^{\infty} 2^{b-1}(b-1)!\ \{tr(\Sigma_2^b) + b\ \underset{\sim}{d}_2'\ \Lambda\ \Sigma_2^b\ \underset{\sim}{d}_2\}\ \frac{w_2^b}{b!} \\ & + \sum_{a=1}^{\infty}\sum_{b=1}^{\infty} 2^{a+b-1}(a+b-2)!\ \{(a+b-1)tr(\Sigma_1^a\Sigma_2^b) + \\ & (a\underset{\sim}{d}_1 + b\underset{\sim}{d}_2)'\ \Lambda\ \Sigma_1^a\Sigma_2^b(a\underset{\sim}{d}_1 + b\underset{\sim}{d}_2) - a\ \underset{\sim}{d}_1'\ \Lambda\ \Sigma_1^a\Sigma_2^b\ \underset{\sim}{d}_1 - \\ & b\ \underset{\sim}{d}_2\ \Lambda\ \Sigma_1^a\Sigma_2^b\ \underset{\sim}{d}_2\}\ \frac{w_1^a}{a!}\ \frac{w_2^b}{b!}\ , \end{aligned} \qquad (C.2.1)$$

where $\underset{\sim}{d}_i = (\underset{\sim}{\mu}-\underset{\sim}{\mu}_i)$, $i = 1,2$, $\Sigma_i = \Lambda^{-1}\Lambda_i$, $i = 1,2$ and w_1, w_2 are some real variables.

Proof : Since $\underset{\sim}{Y}$ is distributed as $N_r(\underset{\sim}{\mu}, \Lambda^{-1})$, the joint cumulant generating function of Q_1 and Q_2 is given by

$$k(w_1,w_2) = \log\left[\{|\Lambda|(2\pi)^{-1}\}^{1/2}\int_{R^r}\exp(w_1Q_1+w_2Q_2)\exp(-\tfrac{1}{2}Q)d\underset{\sim}{y}\right],$$

for some real w_1, w_2, where $Q = (\underset{\sim}{y}-\underset{\sim}{\mu})'\,\Lambda\,(\underset{\sim}{y}-\underset{\sim}{\mu})$.

Let $w_i\,\Lambda_i = V_i$. Then

$$w_1Q_1 + w_2Q_2 - \tfrac{1}{2}Q =$$

$$(\underset{\sim}{y}-\underset{\sim}{\mu}_1)'V_1(\underset{\sim}{y}-\underset{\sim}{\mu}_1) + (\underset{\sim}{y}-\underset{\sim}{\mu}_2)'V_2(\underset{\sim}{y}-\underset{\sim}{\mu}_2) - \tfrac{1}{2}(\underset{\sim}{y}-\underset{\sim}{\mu})'\Lambda(\underset{\sim}{y}-\underset{\sim}{\mu})$$

$$= -\tfrac{1}{2}\underset{\sim}{z}'X\underset{\sim}{z} + 2(\underset{\sim}{d}_1'V_1+\underset{\sim}{d}_2'V_2)\underset{\sim}{z} + \underset{\sim}{d}_1'V_1\underset{\sim}{d}_1+\underset{\sim}{d}_2'V_2\underset{\sim}{d}_2, \qquad (C.2.3)$$

where $\underset{\sim}{z} = \underset{\sim}{y}-\underset{\sim}{\mu}$, $X = \Lambda - 2V_1-2V_2$, and $\underset{\sim}{d}_i = \underset{\sim}{\mu} - \underset{\sim}{\mu}_i$, $i = 1,2$.

This modified expression in (C.2.3) when used in (C.2.2) gives

$$k(w_1,w_2) = -\tfrac{1}{2}\log(\Lambda X^{-1}) + 2(\underset{\sim}{d}_1'V_1+\underset{\sim}{d}_2'V_2)X^{-1}(\underset{\sim}{d}_1'V_1+\underset{\sim}{d}_2'V_2)'$$

$$+ \underset{\sim}{d}_1'V_1\underset{\sim}{d}_1 + \underset{\sim}{d}_2'V_2\underset{\sim}{d}_2. \qquad (C.2.4)$$

If we write $\underset{\sim}{d}_i'V_i\underset{\sim}{d}_i = \underset{\sim}{d}_i'(\Lambda - 2V_1-2V_2)X^{-1}V_i\underset{\sim}{d}_i$, $i = 1,2$

a few terms cancel with each other on the right hand side of (C.2.4), leaving behind

$$k(w_1,w_2) = -\tfrac{1}{2}\log|Z| + \underset{\sim}{d}_1'\Lambda Z^{-1}U_1\underset{\sim}{d}_1 + \underset{\sim}{d}_2'\Lambda Z^{-1}U_2\underset{\sim}{d}_2$$

$$- 2\{(\underset{\sim}{d}_1-\underset{\sim}{d}_2)'\Lambda U_2Z^{-1}U_1(\underset{\sim}{d}_1-\underset{\sim}{d}_2)\}, \qquad (C.2.5)$$

where $Z = (I-2U_1-2U_2)$, $U_i = \Lambda^{-1} V_i$, $i = 1,2$.

We now express all the components of (C.2.5) in the form of power series by means of Lemmas B.1, B.2, and B.3. Firstly, using the result (B.1.1) of Lemma B.1, we can write

$$-\frac{1}{2}\log|Z| = \sum_{a=1}^{\infty} \frac{2^{a-1}}{a} \operatorname{tr}(U_1+U_2)^a.$$

Since $\Lambda = \Lambda_1 + \Lambda_2$ and $\Lambda_1 \Lambda^{-1} \Lambda_2$ is symmetric, we have $U_1U_2 = U_2U_1$. We can, therefore, further expand $(U_1+U_2)^a$ by the result (B.2.1) of Lemma B.2 and get

$$-\frac{1}{2}\log|Z| = \sum_{a=1}^{\infty} \frac{2^{a-1}}{a} \sum_{b=0}^{a} \binom{a}{b} \operatorname{tr}(U_1^a U_2^{a-b})$$

$$= \sum_{a=1}^{\infty} \frac{2^{a-1}}{a} \operatorname{tr}(U_1^a) + \sum_{b=1}^{\infty} \frac{2^{b-1}}{b} \operatorname{tr}(U_2^b) + \sum_{a=1}^{\infty}\sum_{b=1}^{\infty} 2^{a+b-1} \frac{(a+b-1)!}{a!\,b!} \operatorname{tr}(U_1^a U_2^b). \qquad (C.2.6)$$

Further by Lemma B.3, Z^{-1} can be expressed as

$$Z^{-1} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} 2^{i+j} \binom{i+j}{i} U_1^i U_2^j, \qquad (C.2.7)$$

so that

$$d_{\sim 1}' \wedge Z^{-1} U_1 d_{\sim 1} = d_{\sim 1}' \wedge \Big[\sum_{a=1}^{\infty} 2^{a-1} U_1^a \Big\{1 + \sum_{b=1}^{\infty} 2^b \binom{a+b-1}{a-1} U_2^b\Big\}\Big] d_{\sim 1}$$

$$= \sum_{a=1}^{\infty} 2^{a-1}\, d_{\sim 1}' \wedge U_1^a d_{\sim 1} + \sum_{a=1}^{\infty}\sum_{b=1}^{\infty} 2^{a+b-1} \frac{(a+b-1)!}{(a-1)!\,b!}$$

$$d_{\sim 1}' \wedge U_1^a U_2^b d_{\sim 1}. \qquad (C.2.8)$$

Similarly,

$$d_{\sim 2}' \wedge Z^{-1} U_2 d_{\sim 2} = \sum_{b=1}^{\infty} 2^{b-1}\, d_{\sim 2}' \wedge U_2^b d_{\sim 2} + \sum_{a=1}^{\infty}\sum_{b=1}^{\infty} 2^{a+b-1} \frac{(a+b-1)!}{a!\,(b-1)!} d_{\sim 2}' \wedge U_1^a U_2^b d_{\sim 2}. \qquad (C.2.9)$$

Utilizing the same result, i.e., (C.2.7), we also have

$$(\underset{\sim}{d}_1-\underset{\sim}{d}_2)' \wedge U_2 Z^{-1} U_1 (\underset{\sim}{d}_1-\underset{\sim}{d}_2) = \sum_{a=1}^{\infty}\sum_{b=1}^{\infty} 2^{a+b-2} \frac{(a+b-2)!}{(a-1)!(b-1)!} (\underset{\sim}{d}_1-\underset{\sim}{d}_2)' \wedge U_1^a U_2^b (\underset{\sim}{d}_1-\underset{\sim}{d}_2). \tag{C.2.10}$$

Substitution of (C.2.6) and (C.2.8) through (C.2.10) in (C.2.5) and rearrangement of terms further yields

$$\begin{aligned} k(w_1,w_2) = {} & \sum_{a=1}^{\infty} \frac{2^{a-1}}{a} \{tr(U_1^a) + a\underset{\sim}{d}_1' \wedge U_1^a \underset{\sim}{d}_1\} + \\ & \sum_{b=1}^{\infty} \frac{2^{b-1}}{b} \{tr(U_2^b) + b\,\underset{\sim}{d}_2' \wedge U_2^b \underset{\sim}{d}_2\} + \\ & \sum_{a=1}^{\infty}\sum_{b=1}^{\infty} 2^{a+b-1} \frac{(a+b-2)!}{a!\,b!} \{(a+b-1)tr(U_1^a U_2^b) + \\ & (a\underset{\sim}{d}_1+b\underset{\sim}{d}_2)' \wedge U_1^a U_2^b (a\underset{\sim}{d}_1+b\underset{\sim}{d}_2) - a\underset{\sim}{d}_1' \wedge U_1^a U_2^b \underset{\sim}{d}_1 - b\underset{\sim}{d}_2' \wedge U_1^a U_2^b \underset{\sim}{d}_2\} \end{aligned} \tag{C.2.11}$$

Finally, on substituting back $U_i = \Lambda^{-1} V_i = w_i \Lambda^{-1} \Lambda_i$, $i = 1,2$, we obtain the result of (C.2.1), i.e.,

$$\begin{aligned} k(w_1,w_2) = {} & \sum_{a=1}^{\infty} [2^{a-1}(a-1)! \{tr(\Sigma_1^a) + a\underset{\sim}{d}_1' \wedge \Sigma_1^a \underset{\sim}{d}_1\}] \frac{w_1^a}{a!} + \\ & \sum_{b=1}^{\infty} [2^{b-1}(b-1)! \{tr(\Sigma_2^b) + b\,\underset{\sim}{d}_2' \wedge \Sigma_2^b \underset{\sim}{d}_2\}] \frac{w_2^b}{b!} + \\ & \sum_{a=1}^{\infty}\sum_{b=1}^{\infty} [2^{a+b-1}(a+b-2)! \{(a+b-1)\, tr(\Sigma_1^a \Sigma_2^b) + \\ & (a\underset{\sim}{d}_1+b\underset{\sim}{d}_2)' \wedge \Sigma_1^a \Sigma_2^b (a\underset{\sim}{d}_1+b\underset{\sim}{d}_2) - a\underset{\sim}{d}_1' \wedge \Sigma_1^a \Sigma_2^b \underset{\sim}{d}_1 - \\ & b\,\underset{\sim}{d}_2' \wedge \Sigma_1^a \Sigma_2^b \underset{\sim}{d}_2\}] \frac{w_1^a}{a!} \frac{w_2^b}{b!}, \end{aligned}$$

where $\underset{\sim}{d}_i = (\underset{\sim}{\mu} - \underset{\sim}{\mu}_i)$ and $\Sigma_i = \Lambda^{-1}\Lambda_i$, $i = 1,2$.

Remark : The cumulants k_{ao}, k_{ob}, and k_{ab} can be obtained from the function $k(w_1,w_2)$ in (B.5.12) by separating out the coefficients of $(w_1^a/a!)$, $(w_2^b/b!)$, and $(w_1^a w_2^b/a!b!)$, respectively. For example, the joint cumulant k_{ab} is given by

$$k_{ab} = 2^{a+b-1}(a+b-2)!\,\{(a+b-1)\mathrm{tr}(\Sigma_1^a\Sigma_2^b)$$
$$+ (a\underset{\sim}{d}_1+b\underset{\sim}{d}_2)'\,\Lambda\,\Sigma_1^a\Sigma_2^b(a\underset{\sim}{d}_1+b\underset{\sim}{d}_2)$$
$$- a\underset{\sim}{d}_1'\,\Lambda\,\Sigma_1^a\Sigma_2^b\underset{\sim}{d}_1 - b\underset{\sim}{d}_2'\,\Lambda\,\Sigma_1^a\Sigma_2^b\underset{\sim}{d}_2\}. \qquad (C.2.12)$$

A P P E N D I X D

SOME USEFUL RESULTS AND FORMULAE

D.1 Gamma Function :

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} \exp(-\alpha y)\, dy, \quad \alpha > 0$$

D.2 Beta Function in terms of Gamma Function :

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\ \Gamma(\beta)}{\Gamma(\alpha+\beta)} .$$

D.3 Inverse Gamma Function :

$$\Gamma(\alpha) = b^\alpha \int_0^\infty y^{-(\alpha+1)} \exp(-b/y)\, dy, \quad \alpha>0, b>0 \qquad \text{(D.2.1)}$$

D.4 An r-Variate t-distribution [Giri (1977)] :

The r-random vector $\underset{\sim}{Y}$ is said to have an r-vector $\underset{\sim}{Y}$ is distribution, to be designated as $t_r(\underset{\sim}{\mu}, \tilde{\Sigma}, \nu)$, if its p.d.f. is given by

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \frac{\Gamma(\frac{\nu+r}{2})}{\{\Gamma(\frac{1}{2})\}^r \quad \Gamma(\frac{\nu}{2})\ \nu^{r/2}} |\tilde{\Sigma}|^{-1/2} \times$$

$$[1 + (\underset{\sim}{y}-\underset{\sim}{\mu})' \tilde{\Sigma}^{-1} (\underset{\sim}{y}-\underset{\sim}{\mu})]^{-(\nu+r)/2}, \qquad \text{(D.4.1)}$$

where $\tilde{\Sigma}$ is an $r \times r$ symmetric positive definite matrix.

D.5 Inversion Formulae [Cook (1951)] :

The following formulae are appli able to cases $a+b \leq 6$ and given μ_{ab} in terms of k_{ab} for $a \geq b$ only; interchange of suffices will give μ_{ba}

$$\mu_{10} = k_{10} ,$$

$$\mu_{20} = k_{20}+k_{10}^2 ,$$

$$\mu_{11} = k_{11} + k_{10}\, k_{01} ,$$

$$\mu_{30} = k_{30} + 3k_{20}k_{10} + k_{10}^3 ,$$

$$\mu_{21} = k_{21}+k_{20}k_{01} + 2k_{11}k_{10}+k_{10}^2k_{01} ,$$

$$\mu_{40} = k_{40}+4k_{30}k_{10}+3k_{20}^2+6k_{20}k_{10}^2+k_{10}^4 ,$$

$$\mu_{31} = k_{31}+k_{30}k_{01}+3k_{21}k_{10}+3k_{20}k_{11}+3k_{20}k_{10}k_{01}+3k_{11}k_{10}^2+k_{10}^3k_{01},$$

$$\mu_{22} = k_{22}+2k_{21}k_{01}+k_{20}k_{01}^2+k_{20}k_{02}+2k_{12}k_{10}+2k_{11}^2+ 4k_{11}k_{10}k_{01}+k_{10}^2k_{01}^2+k_{10}^2k_{02},$$

$$\mu_{50} = k_{50}+5k_{40}k_{10}+10k_{30}k_{20}+10k_{30}k_{10}^2+15k_{20}^2k_{10}+10k_{20}k_{10}^3+k_{10}^5,$$

$$\mu_{41} = k_{41}+k_{40}k_{01}+4k_{31}k_{10}+4k_{30}k_{11}+4k_{20}k_{10}k_{01}+6k_{21}k_{20}+6k_{21}k_{10}^2$$
$$+3k_{20}^2k_{01}+12k_{20}k_{11}k_{10}+6k_{20}k_{10}^2k_{01}+4k_{11}k_{10}^3+k_{10}^4k_{01},$$

$$\mu_{32} = k_{32}+2k_{31}k_{01}+k_{30}k_{01}^2+k_{30}k_{02}+3k_{22}k_{10}+6k_{21}k_{11}+6k_{21}k_{10}k_{01}$$
$$+3k_{20}k_{12}+6k_{20}k_{11}k_{01}+3k_{20}k_{10}k_{01}^2+3k_{20}k_{10}k_{02}+3k_{12}k_{10}^2$$
$$+ 6k_{11}^2k_{10}+6k_{11}k_{10}^2k_{01}+k_{10}^3k_{01}^2+k_{10}^3k_{02} ,$$

$$\mu_{60} = k_{60}+6k_{50}k_{10}+15k_{40}k_{20}+15k_{40}k_{10}^2+10k_{30}^2+60k_{30}k_{20}k_{10}+$$
$$+ 20k_{30}k_{10}^3+15k_{20}^3+45k_{20}^2k_{10}^2+15k_{20}k_{10}^4+k_{10}^6,$$

$$\mu_{51} = k_{51}+k_{50}k_{01}+5k_{41}k_{10}+5k_{40}k_{11}+5k_{40}k_{10}k_{01}+10k_{31}k_{20}+10k_{31}k_{10}^2$$
$$+ 10k_{30}k_{21}+10k_{30}k_{20}k_{01}+20k_{30}k_{11}k_{10}+10k_{30}k_{10}^2k_{01}+30k_{21}k_{20}k_{10}$$
$$+ 10k_{21}k_{10}^3+15k_{20}^2k_{11}+15k_{20}^2k_{10}k_{01}+30k_{20}k_{11}k_{10}^2+10k_{20}k_{10}^3k_{01}+15k_{11}k_{10}^4+k_{10}^5k_{01},$$

$$\mu_{42} = k_{42}+2k_{41}k_{01}+k_{40}k_{01}^2+k_{40}k_{02}+4k_{32}k_{10}+8k_{31}k_{11}+8k_{31}k_{10}k_{01}$$
$$+ 4k_{30}k_{12}+8k_{30}k_{11}k_{01}+4k_{30}k_{10}k_{01}^2+4k_{30}k_{10}k_{12}+6k_{22}k_{20}$$
$$+ 6k_{22}k_{10}^2+6k_{21}^2+12k_{21}k_{20}k_{01}+24k_{21}k_{11}k_{10}+12k_{21}k_{10}^2k_{01}$$
$$+ 3k_{20}^2k_{01}^2+3k_{20}^2k_{02}+12k_{20}k_{12}k_{10}+12k_{20}k_{11}^2+24k_{20}k_{11}k_{10}k_{01}$$
$$+ 6k_{20}k_{10}^2k_{01}^2+6k_{20}k_{10}^2k_{02}+4k_{12}k_{10}^3+12k_{11}^2k_{10}^2+8k_{11}k_{10}^3k_{01}+k_{10}^4k_{01}^2+k_{10}^4k_{02},$$

$$\mu_{33} = k_{33}+3k_{32}k_{01}+3k_{31}k_{01}^2+3k_{31}k_{02}+k_{30}k_{01}^3+3k_{30}k_{01}k_{02}+k_{30}k_{03}$$
$$+ 3k_{23}k_{10}+9k_{22}k_{11}+9k_{22}k_{10}k_{01}+9k_{21}k_{12}+18k_{21}k_{11}k_{01}+9k_{21}k_{10}k_{01}^2$$
$$+ 9k_{21}k_{10}k_{02}+3k_{20}k_{13}+9k_{20}k_{12}k_{01}+9k_{20}k_{11}k_{01}^2+9k_{20}k_{11}k_{02}$$
$$+ 3k_{20}k_{10}k_{01}^3+9k_{20}k_{10}k_{01}k_{02}+3k_{20}k_{10}k_{03}+3k_{13}k_{10}^2+18k_{12}k_{11}k_{10}$$
$$+ 9k_{12}k_{10}^2k_{01}+6k_{11}^3+18k_{11}^2k_{10}k_{01}+9k_{11}k_{10}^2k_{01}^2+9k_{11}k_{10}^2k_{02}$$
$$+ k_{10}^3k_{01}^3+3k_{10}^3k_{01}k_{02}+k_{10}^3k_{03}.$$

D.6 An Alternative Expression for $[\{(a+r)/4\}\log_e(1+xa^{-1})]$

Using expansion of $\log_e(1 + xa^{-1})$ in terms of (xa^{-1}) we have

$$\left(\frac{a+r}{4}\right)\log_e(1+xa^{-1})$$

$$= \left(\frac{a+r}{4}\right)\left[xa^{-1} - \frac{x^2}{2}a^{-2} + \frac{x^3}{3}a^{-3} - \frac{x^4}{4}a^{-4} + \frac{x^5}{5}a^{-5} - \frac{x^6}{6}a^{-6} + \frac{x^7}{7}a^{-7} + \ldots\right]$$

$$= \left[\frac{x}{4} - \frac{x^2}{8}a^{-1} + \frac{x^3}{12}a^{-2} - \frac{x^4}{16}a^{-3} + \frac{x^5}{20}a^{-4} - \frac{x^6}{24}a^{-5} + \frac{x^7}{28}a^{-6} + \ldots\right]$$

$$+ r\left[\frac{x}{4}a^{-1} - \frac{x^2}{8}a^{-2} + \frac{x^3}{12}a^{-3} - \frac{x^4}{16}a^{-4} + \frac{x^5}{20}a^{-5} + \frac{x^6}{24}a^{-6} + \ldots\right]$$

$$= \frac{x}{4} + \frac{1}{8}(2rx-x^2)a^{-1} + \frac{1}{24}(2x^3-3rx^2)a^{-2} + \frac{1}{48}(4rx^3-3x^4)a^{-3}$$

$$+ \frac{1}{80}(4x^5-5rx^4)a^{-4} + \frac{1}{120}(6rx^5-5x^6)a^{-5} + \frac{1}{168}(7rx^6-6x^7)a^{-6} + \ldots$$

i.e.,

$$\left(\frac{a+r}{4}\right)\log_e(1+xa^{-1}) = \frac{x}{4} + \sum_{i=1}^{\infty} c_i(a^{-1})^i.$$

So far as the coefficients of $(a^{-1})^i$, $i = 1,2,\ldots,\infty$ on the right hand side of the above equation are concerned the following algorithm can be seen to be effective in specifying any coefficient

$$c_i = \frac{(-1)^{i-1}}{4i(i+1)}\left[(i+1)\,rx^i - ix^{(i+1)}\right]$$

$$i = 1,2,\ldots,\infty.$$

REFERENCES

Anderson, T.W. (1957). "An Introduction to Multivariate Statistical Analysis ." Wiley, New York.

Andrews, D.F. (1971). Sequentially designed experiments for screening out bad models with F tests. Biometrika, 58, 427-432.

Atkinson, A.C. (1978). Posterior probabilities for choosing a regression model. Biometrika, 65, 39-48.

Atkinson, A.C. and Cox, D.R. (1974). Planning experiments for discriminating between models. J. Roy. Statist. Soc. B, 36, 321-348.

Atkinson, A.C. and Fedorov, V.V. (1975a). The design of experiments for discriminating between two rival models. Biometrika, 62, 57-70.

Atkinson, A.C. and Fedorov, V.V. (1975b). Optimal design: experiments for discriminating between several models. Biometrika, 62, 289-303.

Bellman, R. (1974). "Introduction to Matrix Analysis ." Tata McGraw Hill Co., New Delhi.

Box, G.E.P. (1949). A general distribution theory for a class of likelihood criteria. Biometrika, 36, 317-346.

Box, G.E.P. and Henson, T.L. (1970). Some aspects of mathematic modelling in chemical engineering. Proc. Inaug. Conf. Scient.Co Centre, Cairo Univ. Press, Cairo.

Box, G.E.P. and Hill, W.J. (1967). Discriminating among mechanistic models. Technometrics, 9, 57-71.

Box, M.J. (1966). A comparison of several current optimization methods, and the use of transformations in constrained problems. Comput. J., 9, 67-77.

Buzzi-Ferraris, G. and Forzatti, P. (1983). A new sequential experimental design procedure for discriminating among rival models. Chem. Eng. Sc., 38, 225-232.

Buzzi-Ferraris, G., Forzatti, P., Emig, G., and Hofman, H. (1984) Sequential experimental design for model discrimination in case of multiple responses. Chem. Eng. Sc., 39, 81-85.

Cook, M.B. (1951). Bi-variate k-statistics and cumulants of their joint sampling distribution. Biometrika, 38, 179-195.

Domez, F. and Froment, G.F. (1976). Dehydrogenation of 1-Butene into Butadiene. Kinetics, catalyst cooking, and reactor design. Ind. Eng. Chem. Proc. Design and Development, 15, 291-301.

Fedorov, V.V. (1972). "Theory of Optimal Experiments." Academic Press., New York.

Fedorov, V.V. and Pazman, A. (1968). Design of physical experiments. Fortschritte der Physik., 16, 325-355.

Froment, G.F. (1975). Model discrimination and parameter estimation in heterogeneous catalysis. AICh.E J., 21, 1041-1057.

Froment, G. and Mezaki, R. (1970). Sequential discrimination and estimation procedures for rate modelling in heterogeneous catalysis. Chem. Eng. Sc., 25, 293-300.

Giri, N.C. (1977). "Multivariate Statistical Inference." Academic Press, New York.

Hellinger, E. (1909). Neue bergrundung der Theorie gradratisher formen Von unendlichvielen Verauderlichen. J. fur. reine and angew. Mathematic, 136, 210.

Hill, P.D.H. (1976). Optimal experimental designs for model discrimination. Ph.D. Thesis, Univ. of Glasgow, U.K.

Hill, P.D.H. (1978). A review of experimental design procedures for regression model discrimination. Technometrics, 20, 15-21.

Hill, W.J. and Hunter, W.G. (1967). Design of experiments for model discrimination in multiresponse situations. Tech. Rep. No. 65, Deptt. of Stat., Univ. of Wisconsin, Madison.

Hill, W.J. and Hunter, W.G. (1969). A note on designs for model discrimination : Variance unknown case. Technometrics, 11, 396-400.

Hill, W.J., Hunter, W.G., and Wichern, D.W. (1968). A joint design criterion for the dual problem of model discrimination and parameter estimation. Technometrics, 10, 145-160.

Hosten, L.H. and Froment, G.F. (1976). Non Bayesian sequential experimental design procedures for optimal discrimination between rival models. Proc. 4th Int. Symp. on Chem. React. Eng. Heidelberg, I1-I13.

Hsiang, T. and Reilly, P.M. (1971). A practical method for discriminating among mechanistic models. Can. J. Chem. Eng., 49, 865-871.

Hunter, W.G. and Mezaki, R. (1967). An experimental design strategy for distinguishing among rival mechanistic models - an application. Can. J. Chem. Eng., 45, 247-249.

Hunter, W.G. and Reiner, A.M. (1965). Designs for discriminating between two rival models. Technometrics, 7, 307-323.

Hunter, W.G. and Wichern, D.W. (1966). Tech. Rep. No. 33, Deptt. Chem. Eng. and Stat., Univ. of Wisconsin, Madison.

Iyengar, S.S. and Rao, M.S. (1983). Statistical techniques in modelling of complex systems : Single and multiresponse models IEEE Trans. on Sys : Man and Cybernatics, SMC-13, 175-189.

Jeffreys, H. (1961). "Theory of Probability." Clarendon Press, Oxford.

Meter, D.A., Pirie, W., and Blot, W. (1970). A comparison of two model discrimination criteria. Technometrics, 12, 457-470.

Pazman, A. and Fedorov, V.V. (1968). Planning of regression and discrimination experiments on NN scattering. Soviet J. of Nuc. Phy., 6, 619-621.

Prasad, K.B.S. and Rao, M.S. (1977). Use of expected likelihood in sequential model discrimination in multiresponse systems. Chem. Eng. Sc., 32, 1411-1418.

Rao, M.S. and Iyengar, S.S. (1984). Application of statistical techniques in modelling of complex systems."Computer Modelling of Complex Biological Systems," CRC Press, Inc., Boca Raton, Florida (Ed. Iyengar, S.S.).

Reilly, P.M. (1970). Statistical methods in model discriminatio Can. J. Chem. Eng., 48, 168-173.

Reilly, P.M. and Blau, G.E. (1974). The use of statistical methods to build mathematical models for chemical racting systems. Can. J. Chem. Eng., 52, 289-299.

Roth, P.M. (1965). Design of experiments for discriminating among rival models. Ph.D. Thesis. Princeton Univ., USA.

Shannon, C.E. (1948). "A mathematical theory of communication." Bell. System. Tech. J., 27, 379-423.

Siddik, S.M. (1972). Kullback-Leibler information function and the sequential selection of experiments to discriminate among several linear models. Ph.D. Thesis, Univ. of Pennsylvania, USA.

Singh, S. and Rao, M.S. (1981). Parameter estimation and model discrimination in multiresponse models. Indian Chem. Engineer, XXIII, 19-26.

Wentzheimer, W.W. (1970). Modelling of heterogeneous catalysed reactions using statistical experimental designs and data analysis. Ph.D. Thesis, Univ. of Pennsylvania, USA.